Strange Floods: The Upper Tail of Flood Peaks in the US

James A. Smith[1], **Alexander A. Cox**[1], **Mary Lynn Baeck**[1], **Long Yang**[1], **Paul Bates**[2]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA
[2]School of Geographical Sciences, University of Bristol, Clifton, Bristol, UK

Key Points:

- The upper tail of flood peaks in the US is examined through analyses of the "upper tail ratio", the peak discharge for the flood of record divided by the sample 10-year flood, for annual flood peak observations from more than 8000 USGS stream gaging stations.
- The distribution of the upper tail ratio does not depend on drainage area, provided that sample size is accounted for, supporting a GEV-based scaling theory in which the upper tail ratio depends only on sample size and the GEV shape parameter.
- Record floods in June 1903, July 1956 and July 1965 in the Blue Mountains of Oregon reflect two of the most important aspects of strange floods in the US: they are located in and adjacent to mountainous terrain and intense thunderstorms are the principal agents of extreme rainfall.
- A fundamental distinction between record floods in the US and the broader population of annual flood peaks is in their seasonal distribution; for record floods, there is a striking shift towards warm season flooding and an increase in occurrence probability during the tropical cyclone season.

Corresponding author: James A. Smith, jsmith@princeton.edu

Abstract

We examine the upper tail of flood peak distributions through analyses of annual peak observations from more than 8000 USGS stream gaging stations and through hydrometeorological analyses of the storms that produce the most extreme floods. We focus on the distribution of the "upper tail ratio", which is defined as the peak discharge for the flood of record at a stream gaging station divided by the sample 10-year flood magnitude. The 14 June 1903 Heppner storm, which produced an upper tail ratio of 200, was the product of a hailstorm that formed along the Blue Mountains in eastern Oregon, a region dominated by snowmelt flooding. A striking contrast between record flood peaks and the larger distribution of annual flood peaks in the US is in the seasonality of flood occurrence, with record floods reflecting a much stronger contribution from warm season thunderstorm systems. Mountainous terrain and intense convective rainfall are important elements of the geography and hydrometeorology of extreme upper tail ratio flood peaks. The distribution of upper tail ratio values for USGS stream gaging stations does not depend on basin area, a result which is consistent with scaling results based on extreme value theory. Downscaling simulations with the Weather Research and Forecasting model are used to examine the storm environment of the 1903 Heppner storm, along with two other record flood peaks near the Blue Mountains of eastern Oregon from the USGS "miscellaneous" flood record, the July 1956 Meyers Canyon flood and the July 1965 Lane Canyon flood.

1 Introduction

The strangest flood in the systematic stream gaging record of the US Geological Survey (USGS) is arguably the 14 June 1903 Heppner flood in eastern Oregon (Figure 1; [*Whistler*, 1903] and [*Byrd*, 2009]). The June 1903 flood peak in Balm Fork (Figure 1) is more than 2 orders of magnitude larger than the 10-year flood peak at the USGS stream gaging station for Balm Fork. The downstream flood peak in Willow Creek at Heppner (Figure 1) is 80 times larger than the 10-year flood for the Willow Creek stream gaging station. The 14 June 1903 Heppner flood, which resulted in more than 250 fatalities, was the product of a hailstorm that formed along the north slope of the Blue Mountains of Oregon ([*Whistler*, 1903]). There are no flood peaks in the systematic stream gaging record of the Blue Mountains region that are comparable to the June 1903 Heppner flood, but there are two "miscellaneous" flood events that have even more extreme estimated flood peaks than the Heppner flood. The July 1956 Meyers Canyon flood ([*Hendricks*, 1964], [*Cooper*, 2006], [*Levish and Ostenaa*, 1996] and [*Costa*, 1987a]) and the July 1965 Lane Canyon flood ([*Rostvedt*, 1970], [*Cooper*, 2006] and [*Costa*, 1987a]) control the US envelope curve ([*Costa*, 1987a]) and the world envelope curve ([*Herschy*, 2003] and [*Costa*, 1987b]) for basin scales between 10 and 40 km^2. In a report on flood frequency for eastern Oregon, *Cooper* [2006] notes that annual peaks for USGS stream gaging stations in eastern Oregon occurred most frequently in spring and resulted from snowmelt. The largest unit discharge flood peaks are "all due to thunderstorms, [but] thunderstorms are essentially unrepresented in the systematic record" ([*Cooper*, 2006]).

We will use the upper tail ratio, which is defined as the peak discharge for the flood of record at a gaging station divided by the sample 10-year flood magnitude, to examine upper tail properties of floods in the US. The 10-year flood is a commonly used threshold to distinguish properties of the upper tail of flood distributions (see, for example, [*O'Connor and Costa*, 2004a] and [*Villarini et al.*, 2011]). In this paper, we will examine strange floods through analyses of USGS flood records from more than 8000 stream gaging stations and through more detailed analyses of the most extreme floods and the storms that produce them.

One of the key notions behind strange floods, as defined by large upper tail ratios at stream gaging stations, is that annual flood peak observations for a drainage basin

may reflect mixtures of different populations of flood agents (see, for example, [*Rossi et al.*, 1984], [*Hirschboeck*, 1987], [*Hirschboeck*, 1988] and [*Smith et al.*, 2011a]), like the snowmelt and thunderstorm floods in eastern Oregon (for similar examples of flood mixutres in Switzerland, see [*Fundel and Zappa*, 2011]). We use the eastern Oregon floods to illustrate the nature of strange floods for which the flood-generating mechanism (orographic thunderstorms) is rare and contrasts markedly with the common flood-generating mechanism (snowmelt). We will show that strange floods can also arise from unprecedented extremes in rainfall for the dominant flood-generating mechanism in a region, as is the case for a collection of record floods in the High Plains and along the Front Range of the Rocky Mountains. Mixtures of flood generating mechanisms can be examined through analyses of the seasonality of flood peaks (see [*Villarini*, 2016] for a recent discussion of seasonality of US flood records). We will show that there is a striking contrast between the seasonality of record flood peaks at USGS stations and the larger record of annual flood peaks from USGS stream gaging stations. The contrast grows even sharper as we restrict attention to record floods with large upper tail ratio values.

The notion of strange floods is also tied to mountainous terrain. Orographic thunderstorms are agents of strange floods in eastern Oregon and in many other settings around the US ([*Costa*, 1987a]) and the world (see, for example [*Herschy*, 2003] and [*Ducrocq et al.*, 2008]). On the extreme end of terrain, Jarrett concluded that 2300 meters is the upper limit for the occurrence of extreme floods in the Front Range of the Rocky Mountains ([*Jarrett and Costa*, 1988] and [*Jarrett and Tomlinson*, 2000]; see also *Mahoney et al.* [2012] for hydrometeorological discussion). Spatial heterogeneity of flood peaks and scaling of flood peaks with drainage area are central topics for characterization of upper tail properties of floods. The literature on flood regionalization and flood peak scaling (see, for example, [*Hosking and Wallis*, 1997], [*Blöschl and Sivapalan*, 1997], [*Robinson and Sivapalan*, 1997], [*Nguyen et al.*, 2013] and [*Ayalew et al.*, 2014]) provides important background for the examination of strange floods.

Hailstorms have been discounted as heavy rainfall producers based on arguments revolving around low precipitation efficiency and rapid storm motion. *Cotton et al.* [2010] note that "storms producing the largest hailstones occur in strongly sheared environments; thus, in general, we should not expect that the storm systems producing the largest hailstones are also heavy rain producing storms" (see also [*Hamada et al.*, 2015], and [*Zipser et al.*, 2006]). *Doswell et al.* [1996] and *Smith et al.* [2001] take a different perspective, pointing to the most intense thunderstorms as underappreciated agents of extreme flooding (see also [*Rogash and Racy*, 2002], [*Hitchens and Brooks*, 2013], [*Nielsen et al.*, 2015], [*Bunkers and Doswell*, 2016], [*Llasat et al.*, 2016] and [*Nielsen and Schumacher*, 2018]).

The 1903, 1956 and 1965 floods in eastern Oregon seem strange partly because we have relatively little insight to the storms that produced them. We utilize downscaling simulations from the Weather Research and Forecasting (WRF) model to reconstruct the environment of the Heppner, Meyers Canyon and Lane Canyon storms and to examine aspects of the structure and evolution of the storms. We have used downscaling simulations to examine the hydrometeorology of the 1927 flooding in the Lower Mississippi River ([*Smith and Baeck*, 2015]) and the 18-19 July 1942 Smethport, Pennsylvania storm that produced world record rainfall at 4-5 hour time scale ([*Smith et al.*, 2011b]; see [*Michaelis and Lackmann*, 2013] for similar analyses).

Measurement error is an important element of flood frequency analysis, especially for the upper tail of flood distributions (see, for example, [*Potter and Walker*, 1981] and [*Costa and Jarrett*, 2008]). The Meyers Canyon and Lane Canyon flood peak estimates have generated considerable controversy ([*Costa and Jarrett*, 2008] and [*Levish and Ostenaa*, 1996]) and reflect the important role that measurement error plays in examining the nature of the upper tail of flood peak distributions.

There are 16 flood events, including the June 1903 Heppner flood, that produced upper tail ratios larger than 20. These provide a catalog of events (Table 1) that we use to examine upper tail properties of floods in the US. Not surprisingly, this catalog includes some of the most devastating floods in US history. Analyses of these events augment and expand analyses based on the entire USGS flood peak record, especially through insights concerning the role of thunderstorms and mountainous terrain in determining upper tail properties of flood peaks.

We introduce a statistical framework based on the Generalized Extreme Value (GEV) distribution to ground our analyses of the upper tail ratio. The GEV has been widely used for flood frequency analysis (see [*Katz et al.*, 2002]), based on arguments from extreme value theory and the representation of the annual maximum flood peak as the maximum of a number of flood peaks over the course of a year. The GEV framework that we develop yields a scaling theory for record floods in which the upper tail ratio does not depend on drainage area. Analyses of upper tail ratios provide insights to upper tail thickness and, more generally, the nature of the upper tail of flood peaks. Robert Horton, along with many of his contemporaries, was convinced that rainfall and floods were bounded, writing "It is not difficult to show from sound meteorological reasoning, and aside from any statistical proof, that there is a natural limitation to rain intensity for any given duration" (letter to F. E. Schmidt, Associate Editor for Engineering News Record; 18 November 1927). We use the upper tail ratio to examine whether bounded flood distributions or unbounded thick-tailed distributions are more consistent with the USGS annual peak record. The literature on envelope curves and record floods (see, for example, [*Enzel et al.*, 1993], [*Vogel et al.*, 2001], [*Douglas and Vogel*, 2006], [*Gaume et al.*, 2009] and [*Halbert et al.*, 2016]) provides important background for these analyses. *Zorzetto et al.* [2016] develop ideas for characterizing the upper tail of hydrologic extremes in which extremes are represented as emergent properties of more common events.

Quantifying the upper tail of flood magnitudes is critical for flood hazard assessment, making strange floods a topic of significant practical importance. Many of these destructive, strange floods led to construction of flood control dams and levees to reduce loss of life and catastrophic damages. Our analyses are designed to advance the understanding of extreme floods and point to new methods for characterizing the upper tail of flood distributions. The US is an important study region both for the societal importance of flood hazards and due to the exceptional observational resources provided by the USGS stream gaging record. Our analyses, however, are designed to provide general insights on extreme flooding and to provide results that can be usefully compared to extreme flooding around the world (see, for example, [*Hosking et al.*, 1985], [*Costa*, 1987b], [*Delrieu et al.*, 2005], [*Gaume et al.*, 2009], [*Parajka et al.*, 2010], [*Li et al.*, 2013], [*Nguyen et al.*, 2013], [*Salinas et al.*, 2014], [*Smith et al.*, 2015], [*Llasat et al.*, 2016] and [*Yang et al.*, 2017]).

The objective of this paper is to characterize the "nature" of the upper tail of flood peak distributions in the conterminous US. To address this objective, we will focus on the following questions: 1) What are the flood-generating mechanisms that control the upper tail of flood distributions in the US? 2) What is the relationship between strange floods and the convective intensity of flood-producing storms? 3) What are characteristic features of watersheds with the largest upper tail ratios in the US? 4) Are strange floods produced by strange storms or by extremes of more common storms? 5) How do upper tail ratio values scale with drainage area and what are the implications for assessing distributional properties of the upper tail of flood peak distributions? 6) Why does eastern Oregon have the strangest floods in the US?

2 Data and Methods

We use USGS stream gaging records, and especially the annual maximum flood peak record (see [*Ryberg et al.*, 2017] for a recent description), to examine strange floods in the conterminous US (i.e., we exclude Hawaii, Alaska and Puerto Rico). Measurements of many extreme floods are made by "indirect discharge" methods, involving field measurements of peak water surface profiles and channel cross-sections, combined with hydraulic computations ([*Costa and Jarrett*, 2008] and [*Koenig et al.*, 2016]). Indirect measurements are made for floods at stream gaging sites when the gage is destroyed or fails to operate properly. They are also made at "miscellaneous" sites, i.e. sites that do not have stream gaging stations, typically for the most extreme floods. The envelope curve of flood peaks for the US is controlled by indirect discharge measurements from miscellaneous sites ([*Costa*, 1987a]). Estimates of the July 1956 Meyers Canyon and July 1965 Lane Canyon flood peaks in eastern Oregon, which are based on indirect discharge methods at miscellaneous sites, have generated considerable controversy (as discussed in Section 4).

The sample of flood peak observations consists of annual maximum instantaneous discharge from USGS stream gaging stations. For a single stream gaging station, let $X_1, ..., X_n$ denote the sample of annual flood peaks and $X_{(1)} < X_{(2)} < ... < X_{(n)}$ denote the order statistics of the annual peak record with n years of observations. With this notation $X_{(n)}$ is the flood of record. For stations that satisfy the stationarity assumption we can represent the flood peak distribution at a site by the cumulative distribution function

$$F(x) = P\{X_i \leq x\} \quad , \quad x > 0 \tag{1}$$

An alternative representation is through the quantile function:

$$Q(p) = F^{-1}(p) \quad , \quad p \in [0, 1] \tag{2}$$

The GEV distribution has been widely used for hydrologic extremes, based on both theoretical and practical considerations ([*El Adlouni et al.*, 2007], [*Martins and Stedinger*, 2000], [*Villarini and Smith*, 2010] and [*Smith et al.*, 2015]). From a theoretical perspective, the GEV represents a limiting distribution that results from the operation of taking the maxima of independent and identically distributed (IID) or weakly dependent random variables (*Leadbetter et al.* 1982). The quantile function of the GEV takes the form:

$$Q(p) = \mu - \sigma \{\frac{(1 - [-ln(p)]^{-\xi})}{\xi}\} \quad , \quad \xi \neq 0 \tag{3}$$

$$= \mu - \sigma \, ln\{-ln(p)\} \quad , \quad \xi = 0 \tag{4}$$

where $\mu \in (-\infty, +\infty)$ is the location parameter, $\sigma \in (0, +\infty)$ is the scale parameter, and $\xi \in (-\infty, +\infty)$ is the shape parameter. For $\xi > 0$, the distribution is unbounded above. For $\xi < 0$, the distribution is bounded above with an upper bound of $\mu - \sigma/\xi$. The Gumbel distribution is the special case for $\xi = 0$ and corresponds to unbounded, but "thin" upper tails. The GEV provides a useful tool for examining tail properties of flood peak distributions. We will use the shape parameter ξ as an index of tail properties (see, for example, [*Morrison and Smith*, 2002]).

We examine the stationarity of annual maximum peak discharge data using the Mann-Kendall test (see [*Helsel and Hirsch*, 1993]). The test has been widely used for hydrologic time series, including annual peak records (see, for example, [*Villarini et al.*, 2009]).

Other methods, including change point analyses, can be used to identify the nature and timing of nonstationarities (see, for example, [*Villarini et al.*, 2009] and [*Helsel and Hirsch*, 1993]). In this study, we will focus on the presence or absence of nonstationarities in a flood record and restrict analyses to those based on the Mann-Kendall test.

We estimated the GEV parameters for stationary annual peak records (5546 USGS stations) using the maximum likelihood method (see [*Coles*, 2001]). The estimated location and scale paramters $\{\hat{\mu}_i\ , \hat{\sigma}_i\}$ are linearly related (correlation of 0.91), with $\hat{\mu}\ = c \times \hat{\sigma}$ and c approximately equal to 2. The estimated location and scale parameters also have significant correlation with drainage area (0.57 for both). The estimated shape parameters $\{\hat{\xi}_i\}$ are independent of the estimated location and scale parameters and drainage area. We examined dependence of the estimated shape parameter on a range of other covariates contained in the USGS Gages-II data base ([*Falcone et al.*, 2010]) and found no significant results.

The upper tail ratio, Z, is the ratio of the flood of record to the sample 10-year flood, i.e.

$$Z\ =\ \frac{X_{(n)}}{\hat{Q}(.9)} \tag{5}$$

where $\hat{Q}(.9)$ is computed in the standard way as a linear combination of the order statistics $X_{(j)}$ and $X_{(j+1)}$, with $\frac{j}{n+1}\ <\ 0.9\ \leq \frac{j+1}{n+1}$. We will use the upper tail ratio as a principal tool for examining the upper tail of all USGS flood records. For nonstationary records, $\hat{Q}(.9)$ can not be interpreted as a 10-year flood for any particular time period, but provides an aggregate characterization of the 1 in 10 probability of occurrence over the changing flood regimes represented in a particular record of a stream gaging station. Of particular importance for our analyses will be stations for which extreme floods provided the stimulus for river regulation, stream gaging or both.

The upper tail ratio provides a nonparametric approach to analyze the upper tail of flood peak distribuitons. Parametric assumptions on the parent flood peak distribuitons from extreme value theory provide guidance on the distribution of the upper tail ratio for stationary flood records. Under the assumptions that: 1) the parent flood peak distribution for a stream gaging station has a GEV distribution and the shape parameter is linearly related to the location parameter, with $\mu\ =\ c\ \sigma$ (see discussion above) and 2) the sample 10-year flood is close to the theoretical 10-year flood, then the upper tail ratio has a GEV distribution that depends only on the shape parameter and sample size (details are provided in the Appendix). This result provides a characterization of the control of the shape parameter and sample size on upper tail properties of the GEV distribution (see [*Koutsoyiannis*, 1999] for related scaling analyses of extreme precipitation). Under the assumption that the GEV provides a good representation of annual flood peaks, the result shows that upper tail ratio and the shape parameter of the GEV are intricately linked. Because the shape parameter does not depend on basin scale, it follows that the same holds for the upper tail ratio.

The form of the GEV distribution for Z does not depend on the selection of the 10-year flood as the normalizing variable. The same form of the distribution (Equation A.3 in the Appendix) holds for any quantile Q(p), with 0.9 replaced by the new value of p (Equation A.4 in the Appendix). More generally, we could define Z as the ratio of the record flood to the sample T-year flood for values of T different from 10 years. Selecting shorter return intervals could provide sample T-year floods with lower variability, but at the cost of reduced relevance to the upper tail of flood distributions. Our selection of the 10-year flood is motivated in part by DuMouchel's admonition to "let the tails speak for themselves" ([*DuMouchel*, 1983]).

3 Strange Floods - The USGS Stream Gaging Record

3.1 Upper Tail Ratio for USGS Stream Gaging Stations

We examine the estimated upper tail ratio for 8911 stream gaging stations that are included in the USGS GAGES-II data set ([*Falcone et al.*, 2010]). For each stream gaging station we have a vector of covariates which includes the number of annual peak observations n_i, drainage area A_i, the time $T_i \in [0, 1]$ of the record flood (with time 0 denoting the beginning of the day on January 1 and time 1 denoting the end of the day on December 31), the mean annual precipitation, mean basin elevation and a wide range of land use / land cover and basin morphology variables taken from the Gages II data set. We principally use time T_i to assess seasonality of flood occurrences. Synoptic analyses of USGS flood records have assessed flood frequency for the US from diverse perspectives (see, for example, [*Patton and Baker*, 1976], [*Michaud et al.*, 2001], [*O'Connor and Costa*, 2004a], [*O'Connor and Costa*, 2004b], [*O'Connor et al.*, 2002], [*Costa*, 1987a], [*Costa*, 1987b], and [*Smith and Smith*, 2015]).

The locations of flood peaks with upper tail ratios greater than 5, 10 and 20 are concentrated in diverse regions across the US (Figure 2), but the largest upper tail ratio floods are found in the western US (Table 1 and Figure 2). The spatial distribution of large upper tail ratio flood peaks reflects many of the features identified by *O'Connor and Costa* [2004a] in their examination of large unit discharge flood peaks. Analyses based on unit discharge flood peaks place a stronger spotlight on somewhat different regions (notably Texas and the Central Appalachians; see [*O'Connor and Costa*, 2004a]) than analyses based on upper tail ratio (see additional discussion below).

An important distributional property of the upper tail ratio is the dependence on record length (Figure 3). The median and .95 quantile values of upper tail ratio for the second decile (annual peak records of 14 - 22 years) are 1.42 and 3.94. For the ninth decile (75 - 86 years of annual peak records) the median and .95 quantile values are 2.13 and 6.37. Dependence of upper tail ratio on sample size is most pronounced for stations in the lower half of the distribution of sample size. For the upper half, i.e. stations with more than 46 years of record, there is only a modest change in the distribution of upper tail ratio with increasing record length.

The GEV scaling theory for upper tail ratio implies that the distribution of upper tail ratio values should depend on the GEV shape parameter, in addition to sample size. This is supported by analyses based on estimated values of the GEV shape parameter (Figure 4). For the analyses in Figure 4 we use stations with sample size greater than 46 (the median value of sample size) to focus on the dependence of upper tail ratio on the shape parameter and to mitigate the impacts of samping variability for estimators of the shape parameter. Median, interquartile range and maximum values of upper tail ratio increase systematically with deciles of estimated shape parameter. Variability of upper tail ratio values increases sharply for the two largest deciles of estimated GEV shape parameters (the lower bound of the estimated shape paramter for the 9th decile is 0.5) which is consistent with prior studies of varaibility of maximum likelihood estimators of the shape parameter (see [*Morrison and Smith*, 2002]).

For the entire sample of 8,911 stream gaging stations, the distribution of upper tail ratio magnitudes increases with basin scale, as represented by drainage area, A_i (Figure 5 top). This result appears to be at odds with the GEV scaling theory for which the distribution of upper tail ratio depends only on the shape parameter and sample size. As noted in the previous section, the estimated shape parameter for stationary flood records does not depend on drainage area, or any other basin characteristic. There is, however, a modest positive correlation between basin area and sample size (figure not shown), which is sufficient to skew the distribution of upper tail ratio. If we restrict consideration to stations with more than 46 years of record (the median value of record length), the de-

pendence of upper tail ratio on basin area disappears (Figure 5 bottom). The distribution of upper tail ratio magnitudes is uniform across the range of basin scales, consistent with the GEV scaling theory.

Record floods, and especially the most extreme record floods, are concentrated during the peak of warm season convection (Figure 6). The distribution of record flood peaks has a maximum around mid-June, with a secondary maximum around the beginning of September, corresponding to the peak of extreme flooding in the eastern US from landfalling tropical cyclones. The distribution sharpens when we restrict consideration to stations with upper tail ratio greater than 5, with the summer maximum becoming more prominent. The tropical cyclone peak shifts earlier in the season, due largely to a handfull of hurricanes (Diane in August of 1955, Alberto in July of 1994 and Agnes in June of 1972), which have a disproportionate impact on the seasonal distribution of upper tail ratios greater than 5.

The contrasts in seasonal distribution of record floods and record floods with upper tail ratio greater than 5 are small, compared with the broad seasonal distribution of annual flood peaks at USGS stream gaging stations (Figure 6), which has winter/spring and summer maxima and a diminished contribution from tropical cyclones. The seasonal distribution of extreme floods becomes even more skewed towards warm season convection and tropical cyclones when attention is restricted to small watersheds (less than 200 km^2 in drainage area; figure not shown).

We also examine strange floods through the record flood seasonality, S_i, which characterizes the seasonal anomaly of the record flood for station i:

$$S_i \;=\; \hat{f}_i(T_i) \tag{6}$$

where $\hat{f}_i(x)$, x $\in$ $[0, 1]$, is the sample probability density function of the time of occurrence of flood peaks in basin i. We compute S_i as the sample probability density function of the seasonal occurrence of annual flood peaks at station i, evaluated at the time of occurrence T_i of the flood of record. We estimate the seasonal probability density function using the R Kernel Density Estimator ([*Venables and Ripley*, 2002]). If the times of occurrence are uniformly distributed over the year, the function $f_i(x)$ takes the value 1. Values of S_i less than 1 indicate record floods that have anomalous components of seasonality. Values that are significantly greater than 1 indicate record floods that are produced by common flood generating mechanisms. Record flood seasonality between stations with upper tail index greater than 5 versus those with upper tail index less than 5 differ only on the high end of record flood seasonality. For upper tail index less than 5, the median value of S_i is 1.43, compared with 1.46 for upper tail index greater than 5. The .95 quantile of record flood seasonality is 7.83 for upper tail index less than 5 and 4.93 for upper tail index greater than 5, indicating a larger percentage of basins with very strong seasonality for upper tail index less than 5. These results reflect the prominence of winter/spring flood peaks, including record floods, for many stream gaging stations in the eastern US.

Arid and semi-arid regions have upper tail ratios that are somewhat larger than those in more humid regions (Figure 7). Upper tail ratios for the lowest two deciles of mean annual basin precipitation are systematically larger than those for basins with greater precipitation. The upper bound of the second decile of precipitation is approximately 650 mm. There is relatively little variation in the distribution of upper tail ratio with precipitation for the upper 8 deciles. As highlighted in Figures 2 and 7, arid/semi-arid regions play an important role in controlling the extremes of upper tail ratio.

Jarrett's conjecture ([*Jarrett and Costa*, 1988] and [*Jarrett and Tomlinson*, 2000]) that 2300 meters elevation marks a significant break in extreme flood records is supported

by analyses of the upper tail ratio for the entire USGS annual peak record (Figure 8). There is a pronouced decrease in median, interquartile range and extremes of upper tail ratio for the last decile of basin elevations, which has a lower bound at 2300 meters. The sharp decrease in upper tail ratio basins with elevations greater than 2300 meters is paired with a general increase in upper tail ratio for deciles 6 - 9 (elevations ranging from approximately 395 meters to 2300 meters). This feature reflects the role of orographic precipitation mechanisms in amplifying flood peak extremes.

For more than a century there has been debate over the impact of forests on floods (see, for example, [*Jones and Grant*, 1996]). There is little evidence of forest cover impacts on upper tail ratio (Figure 9), suggesting that forest cover does not have a major impact on the most extreme floods. This result is consistent with previous studies showing that the impacts of forest cover on flooding diminishes in the upper tail of flood events ([*Wood et al.*, 1990]). These studies point to the relative importance of soil moisture decreasing as return interval (and storm total rainfall accumulations) becomes large. There are caveats to the inferences concerning forests and floods. The Gages II forest cover data represent current conditions and do not capture changes over time in forest cover. There may also be impacts on more common floods that are not present for the most extreme events. For the smallest decile of forest cover, which includes urbanized watersheds and arid watersheds, there is an elevated distribution of upper tail ratio.

Upper tail ratio decreases as impervious fraction increases, implying that increases in urbanization result in lower values of upper tail ratio. This result is consistent with increases in frequency of flooding (which leads to elevated values of 10-year flood) and capacity constraints on peak flood magnitude imposed by the urban drainage network ([*Smith and Smith*, 2015]). Capacity constraints in urban watersheds result from design of components of the urban drainage network, including pipe capacities in the storm drain network and bridge culverts. The decreases in upper tail ratio with impervious fraction do not reflect the nature of flood hazards, which increase in urban regions due to the vulnerability of large flood-impacted populations (see [*DiBaldassarre et al.*, 2013] for discussion and review of human-flood interactions).

The pronounced nonstationarities in flood records of the US (see, for example, [*Milly et al.*, 2008], [*Villarini et al.*, 2009], [*Hirsch and Archfield*, 2015] and [*Mallakpour and Villarini*, 2015]) do not translate into large differences in upper tail ratio. We examine the distributions of upper tail ratio for stations that pass the Mann-Kendall test for stationarity and those that do not. The median value for the nonstationary stations, 1.99, is only slightly larger than the median value for stationary stations, 1.88. There are similarly modest changes for interquartile range and extremes. River regulation by dams is an important element of nonstationarities in flood records. There is a link between large upper tail ratio floods and dam building, as we illustrate in the following section. The occurrence of a large flood in a particular drainage basin marks that watershed as a target for flood control, especially if there is extensive loss of life or damage from the flood.

The sample properties of upper tail ratio for USGS stream gaging stations are most consistent with a GEV world in which the shape parameter is positive (Figures 10 and 11) implying unbounded, thick-tailed flood distributions. In Figure 10 we illustrate the dependence of the distribution of upper tail ratio on record length for GEV distributions with shape parameters of 0.25 (unbounded and thick-tailed), 0 (Gumbel) and -0.25 (bounded), with sample sizes taken from the 8911 USGS stations. The bounded case produces upper tail ratios that are too small at all quantiles and the extremes in the Gumbel case are too small. Only the unbounded, thick-tailed case can produce the extremes that are present in the USGS upper tail ratio samples. In Figure 11 we show the distribution of upper tail ratio for 8000 simulated flood records of length 60 years with shape parameter ranging from -0.5 to 0.5. Again we conclude that bounded and Gumbel distributions are not consistent with the large upper tail ratios for USGS gaging stations. Large upper tail ratios in the GEV world are tied to unbounded thick-tailed distributions. We

have also explored mixtures of GEV families and come to a similar conclusion (figures not shown). These results, along with relationships between estimated GEV shape parameters and upper tail ratio (Figure 4), point to thick-tailed distributions as a key element of large upper tail ratio flood records. Future studies will examine these issues in more detail.

3.2 The Largest Upper Tail Ratio Floods in the US

In this section, we examine the largest upper tail ratio flood peaks in the USGS stream gaging record (Table 1 and Figure 2) to expand the interpretation of results presented in the previous section. The 14 June 1903 flood peak in Balm Fork (Table 1 and Figure 1) translates to an upper tail ratio of 198, the second largest in the USGS stream gaging record (Table 1). Balm Fork enters Willow Creek immediately upstream of Heppner, Oregon. The Willow Creek flood peak from the 1903 flood ranks third, with an upper tail ratio of 80 (Table 1 and Figure 1). The flood peaks for Balm Fork and Willow Creek in 1903 were determined from indirect discharge estimates (see Section 2). Years after the Heppner flood, USGS stream gaging stations were installed in Willow Creek (1949) and Balm Fork (1983), providing the observations to assess the distribution of the upper tail ratio; the record length for the Willow Creek station is 67 years and the record length for the Balm Fork station is 33 years. The Heppner flood is the quintessential strange flood. It has the largest upper tail ratio for a meteorological flood in the US and it was produced by an unusual flood agent for the region, i.e. a hailstorm in complex terrain. As *Cooper* [2006] notes, flood peaks in the Blue Mountains region of eastern Oregon are largely from snowmelt.

The largest upper tail ratio in the USGS flood record, 319 from the 5 June 1977 flood peak of 48140 $m^3\ s^{-1}$ on the Teton River near St. Anthony, Idaho (2295 km^2), highlights the extreme nature of the June 1903 Heppner flood. The Teton River flood resulted from the failure of the Teton Dam, which was located a short distance upstream from the gaging station. One of the most notable dam failures in US history (see [*O'Connor et al.*, 2002] for discussion of dam-break floods), the resulting flood devastated downstream communities. The value of the Balm Fork upper tail ratio is closer to the Teton River upper tail ratio than it is to the Willow Creek upper tail ratio (Table 1).

The 31 May 1935 Republican River Flood, which produced an upper tail ratio of 71 for the McCook, Nebraska gaging station (6838 $m^3\ s^{-1}$ peak; see Table 1), is notable for the relatively large drainage area, 31700 km^2, and for the combined occurrence of extreme flooding and severe weather ([*Follansbee and Spiegel*, 1937]). In addition to the upper tail ratio of 71 at the McCook gaging station, 4 other USGS stream gaging stations reported record flood peaks with upper tail ratios greater than 20 (Table 1).

The Republican River basin is located in a corridor of high frequency occurrence of extreme rainfall events during the May - July period. The corridor extends from the Front Range of the Rocky Mountains through Nebraska into Iowa. This feature is linked to the climatological maximum of rainfall from mesoscale convective complexes (MCCs; see [*Maddox*, 1980], [*McAnelly and Cotton*, 1989], [*Tuttle and Davis*, 2006] and [*Zhang et al.*, 2001]), and more generally, the climatology of mesoscale convective systems ([*Schumacher and Johnson*, 2005]). Heavy rainfall on 30 May 1935 initiated along the eastern margin of the Rocky Mountains near Colorado Springs, Colorado around the middle of the day and the path of heavy rainfall extended northeast through eastern Colorado, northwestern Kansas and into Nebraska on May 31; peak rainfall accumulations exceeded 500 mm. The record flood seasonality index, S_i for the Republican River flood of 3.0 (Table 1) is large, reflecting the prominence of mesoscale convective systems in the flood record for the region.

Like the Heppner flood, the principal flood agents for the Republican River flooding were from the most intense spectrum of thunderstorms. Multiple bucket survey re-

ports noted that precipitation accumulations included rain and hail ([*Follansbee and Spiegel*, 1937]). As storms passed McCook, Nebraska, a tornado was reported ([*Follansbee and Spiegel*, 1937]). Unlike the Heppner flood, the 1935 Republican River flood is not strange in terms of the flood agent at play, but in terms of the magnitudes of rainfall and flood peaks. Extreme rainfall and flooding extended down the High Plains from Nebraska to Texas. The world record rainfall at 2.75 hour time scale is the 559 mm rainfall accumulation in D'Hannis, Texas, on 31 May 1935. The D'Hannis storm was a tornadic supercell thunderstorm ([*Smith et al.*, 2000]) that produced an envelope curve flood peak in Seco Creek (based on indirect discharge measurements; see [*Costa*, 1987a] and [*Costa and Jarrett*, 2008]).

A common consequence of the largest floods in the USGS flood record has been to stimulate development of flood control systems, typically through dam building. The May-June 1935 flooding was a major impetus for the 1944 Pick-Sloan Plan ([*Billington and Jackson*, 2006]), which resulted in a system of dams in the Republican River basin, and the Flood Control Act of 1936, which accelerated dam building throughout the US. The annual peak records for the Republican River stream gaging stations all exhibit non-stationarities and these are tied to dam building after implementation of the Pick-Sloan Plan.

The High Plains region of the central US is an area with a concentration of large upper tail ratios extending from Texas to North Dakota (Figure 2 and Table 1). Like the Republican River flood of 1935, these events are dominated by mesoscale convective systems. The Washita flood of 3 April 1934 in Oklahoma and Prairie Dog Creek flood of 28 May 1953 in Kansas (listed in the USGS annual peak records as 23 May 1953) are two of the most extreme (Table 1). The Washita flood, which ranks just behind the Republican River flood in upper tail ratio (Table 1), resulted in 17 fatalities and stimulated development of a flood control system for the Washita River basin. Newspaper accounts of the storm refer to the frequent occurrence of lightning. The Prairie Dog Creek flood had an upper tail ratio of 37 (7th on the USGS list; Table 1) and was produced by rains of 8 to 11 inches, which resulted in a "phenomenal rise in Prairie Dog Creek" ([*Wells*, 1959]). Extreme rainfall for the Prairie Dog Creek storm was associated with a tornadic thunderstorm (based on NCDC Storm Reports).

The most striking expression of mountainous terrain on the occurrence of large upper tail ratio floods is in the Front Range of the Rocky Mountains (Figure 2; see [*Follansbee and Jones*, 1922], [*Maddox et al.*, 1978], [*Jarrett and Costa*, 1988], [*Ogden et al.*, 2000], [*Petersen et al.*, 1999], [*Javier et al.*, 2007], [*Mahoney et al.*, 2012]). By many metrics, including the upper tail ratio (Table 1), the most extreme flooding in the Front Range region resulted from a series of storms during the period from 16-18 June 1965 (in terms of fatalities, however, the 21 deaths in 1965 rank well behind the more than 140 fatalities in Big Thompson on 31 July 1976 and the "hundreds" of fatalities in Pueblo on 4 June 1921; see [*Follansbee and Jones*, 1922] and [*England et al.*, 2014]). The Plum Creek flood peak in suburban Denver on 16 June 1965 resulted in an upper tail ratio of 55 at a drainage area of 783 km^2 (and, not surprisingly, development of flood control for Plum Creek, Bear Creek and the South Platte River). The Jimmy Camp Creek flood peak the following day produced an upper tail ratio of 34 at a drainage area of 170 km^2 and an envelope curve flood peak for the US ([*Costa*, 1987b]). The storms were severe thunderstorms, especially the Jimmy Camp Creek storms, which were associated with large hail and F1 tornadoes. The only overlap between Costa's US flood peak envelope curve ([*Costa*, 1987a]) and the strange floods in the systematic record (Table 1) is the Jimmy Camp Creek flood.

The 16-18 June 1965 period was a major flood outbreak along the Front Range from Colorado south into New Mexico. The Cimarron River near Cimarron, New Mexico on June 17 (708 km^2) reported a record flood with an upper tail ratio of 27. Rayado Creek near Cimarron, New Mexico (159 km^2) reported a record flood peak with an upper tail

ratio of 21. Tornado reports (NCDC Storm Events data base) suggest that the Rayado Creek flood was produced by extreme rainfall from a supercell thunderstorm. The record flood seasonality for the Front Range watersheds experiencing record floods during the 16-18 June 1965 period range from 1.5 to 2.2 (Table 1) reflecting the prominence of warm season thunderstorms in the flood records of the Front Range region. The June 1965 flood was not strange in terms of the flood-generating agent, but in the magnitude of rainfall.

The combination of mountainous terrain and organized thunderstorms resulted in one of the most deadly floods in US history in Rapid City, South Dakota on 9 June 1972. The Rapid City flood (also known as the Black Hills Flood) was not just one of the most extreme floods in US history in terms of upper tail ratio (49 at a drainage area of 973 km^2, with 5 stations exceeding an upper tail ratio of 20; see Table 1), but also in terms of fatalities, with at least 238 deaths that were the direct result of the flood. The flood was produced by a nearly stationary system of thunderstorms, with orographic mechanisms playing an important role in storm structure and evolution ([*Maddox et al.*, 1978]; see also [*Soderholm et al.*, 2014]). The Black Hills region has an even more pronounced seasonal concentration of flood peaks in June than the Front Range, with record flood seasonality for the five stations ranging from 3.7 - 5.0 (Table 1). Paleoflood studies ([*Harden et al.*, 2011]) have shown that the June 1972 floods were not unprecedented; during the past 2000 years multiple flood peaks as large or larger than the June 1972 flood occurred in the watersheds most heavily impacted by the June 1972 storms.

Although Jarrett's 2300 meter hypothesis is supported by results in Figure 8, there are exceptions to the 2300 meter rule, with one ranking among the largest upper tail ratio flood events. The Costilla Creek flood of 22 July 1954 (Table 1) occurred in a watershed with an elevation of 2880 meters at the basin outlet. Costilla Creek is located in the Sangre de Christo Mountains of northern New Mexico. The 22 July 1954 flood peak occurred during the North American Monsoon season. Snowmelt is the dominant flood agent in the watershed, with 70% of annual flood peaks occurring during the April - June snowmelt season. The third and fourth largest peaks, however, are North American Monsoon events occurring in August 1957 and August 1982. The record flood seasonality of 2.1 for the 22 July 1954 peak (Table 1) reflects the bimodal distribution of flood seasonality in the watershed. Snowmelt dominates, but there is a significant contribution to the flood record from North American Monsoon thunderstorms. The magnitude of the 22 July 1954 flood peak is unprecedented, not the occurrence of a flood during the North American Monsoon season.

The Sybille Creek flood of 20 August 1990 in Wyoming, which produced an upper tail ratio of 35 (Table 1), reprises themes from the Heppner flood (see G. F. Ritz's summary in [*Jordan and Combs*, 1996]). The Sybille Creek flood, like the Heppner flood and many of the large upper tail ratio events, was the product of an intense hailstorm in which extreme rainfall and exceptional hail falls were in close proximity. Accumulations of hail reached several feet in portions of the Sybille Creek watershed.

The Teton River of Montana was the setting for a strange hydrologic flood, with a record peak on 8 June 1964 producing an upper tail ratio of 21 (Table 1; [*Boner and Stermitz*, 1967]). Although geographically close to the Teton River of Idaho (Table 1), the Montana river is tributary to the Missouri and Mississippi, flowing to the Gulf of Mexico; the Idaho Teton River is a tributary to the Snake and Columbia, flowing to the Pacific Ocean. In a supplement to *Boner and Stermitz* [1967], R. A. Dightman notes that "an item of more than passing interest was the lack of thunderstorms in the heavy precipitation region". This observation reflects the view of the hydrometerological community in the US Weather Bureau and the flood hydrology community of the USGS from the 1930s starting with Robert Follansbee and Nathan Grover through the 1960s that thunderstorms play an important role as agents of extreme flooding. In the 1920s and 1930s, USGS hydrologists documented extreme rainfall and flood runoff from "cloudburst" thunderstorms (see, for example, [*Follansbee and Jones*, 1922] and [*Woolley*, 1946]). The

USGS view of floods during the 1950s and 1960s was clearly influenced by Luna Leopold (Chief Hydrologist from 1956 - 1966). Leopold's background in meteorology and his early research on "cloudburst" floods in the southwestern US ([*Leopold*, 1942] and [*Leopold*, 1946]) gave him a deep appreciation for the role of thunderstorms as flood agents in the US.

The Ashland Creek floods in the mountains of southern Oregon (Figure 2 and Table 1) on 15 January 1974 reflect a prominent setting for extreme flooding in northern California and the Pacific Northwest. Heavy orographic rainfall from extratropical systems, combined with rapid snowmelt associated with strong warm air advection ([*Colle and Mass*, 2000]), can produce exceptionally large flood peaks in the region. These events are linked to concentrated low-level transport of water vapor in "atmospheric rivers" (in contrast to the Blue Mountain floods in eastern Oregon, as discussed below).

A cluster of stations with upper tail ratio greater than 5 is located in southern California along the San Gabriel Mountains north of Los Angeles (Figure 2). This cluster is dominated by the 3-4 March 1938 flood for which 9 stations reported record floods with upper tail ratios greater than 5; the San Antonio Creek station has an upper tail ratio of 24 (Table 1). The flood devastated portions of Los Angeles County and resulted in more than 100 fatalities. The March 1938 storms provide additional insights to the 2300 meter hypothesis for extreme rainfall and flooding. Rainfall at Kelly's Kamp at an altitude of 2530 meters in the headwaters of San Antonio Creek was 810 mm during the period from 27 February to March 4. The March 1938 flood, like the January 1974 flood in southern Oregon and many of the major floods along the west coast of the US was produced by major Pacific storms during the winter season (see, for example, [*Dettinger et al.*, 2011], [*Neiman et al.*, 2011] and [*Barth et al.*, 2017]).

The San Emigdio Creek flood on 5 August 1961 along the north slope of the San Emigdio Mountains north of Los Angeles, California (Table 1) is both strange and mysterious. The seasonal distribution of flood peaks in San Emigdio Creek is dominated by winter storms, resulting in the smallest value of the record flood seasonality S_i, 0.66, among the events with upper tail ratio greater than 20. Exceptional floods from summer cloudbursts in San Emigdio Creek have, however, been described by *Douglas* [1908], who notes that a high-gradient tributary is named Cloudburst Canyon.

Texas has a well-recognized history of extreme floods and it is a region with a high concentration of large upper tail ratio floods (Figure 2) and a diverse population of flood mechanisms ([*Baker*, 1975], [*Patton and Baker*, 1976], [*Smith et al.*, 2000], [*O'Connor and Costa*, 2004a], [*Nielsen-Gammon et al.*, 2005] and [*Zhu and Quiring*, 2013]), including severe thunderstorms as noted above for the May 1935 Seco Creek flood. The Medio Creek flood of 22 September 1967 is the only Texas flood and the only tropical cyclone flood to produce an upper tail ratio greater than 20 (Table 1). Hurricane Beulah produced extreme rainfall in south Texas, with a maximum accumuation of 695 mm in 36 hours in Pettus, Texas, which is located in the upper portion of the Medio Creek watershed. More extreme rainfall accumulations have been recorded in Texas. The 24-h record rainfall for the United States of 1090 mm occurred at Alvin, Texas on 25-26 July 1979 (Tropical Storm Claudette), returning to southeastern Texas the record that was held for many years by the 965 mm at Thrall, Texas on 10 September 1921 (Hurricane 1 of 1921) near the Balcones Escarpment ([*Smith et al.*, 2000] and [*Caracena and Fritsch*, 1978]). Hurricane Harvey produced 5-day accumulations (25 - 29 August 2017) of 1250 mm in Houston, the largest for a single storm in the conterminous US (see *Emanuel* [2017] for climatological assessments of tropical cyclone rainfall frequency in Texas).

Floods with large upper tail ratio in the US east of the Mississippi River are dominated by tropical cyclones (Figure 2). Five hurricanes account for more than 40% of the 123 stations east of 90.2 degrees west with upper tail ratios greater than 5. Hurricane Diane in August 1955 accounted for 19 of the stations, the Great New England Hurri-

cane of September 1938 acounted for 9, Hurricane Floyd in September 1999 and Hurricane Alberto in July 1994 accounted for 8 and Hurricane Agnes in June 1972 accounted for 7. Tropical cyclones account for 2/3 of the 123 record flood events east of the Mississippi River with upper tail ratio greater than 5.

There is large diversity in rainfall and flooding from this collection of tropical cyclones. Diane produced abnormally large rainfall rates and retained tropical characteristics throughout its passage through the northeastern US. Most tropical cyclones that reach the mid-Atlantic and New England experience extratropical transition through interaction with mid-latitude extratropical systems ([*Hart and Evans*, 2001] and [*Atallah and Bosart*, 2003]). Hurricanes Agnes ([*DiMego and Bosart*, 1982]), Floyd ([*Colle*, 2003] and [*Atallah and Bosart*, 2003]) and Irene ([*Liu and Smith*, 2016]) reflect rainfall distributions characteristic of extratropical transition events, producing extensive regions of heavy rainfall. For Hurricane Agnes, there were 125 stations reporting record floods, compared with 68 stations from Hurricane Diane.

Orographic thunderstorms in the central Appalachians provide another class of strange floods ([*Giordano and Fritsch*, 1991], [*Konrad*, 2001], [*Smith et al.*, 2011b]). *Villarini and Smith* [2010] show that orographic thunderstorms are largely unrepresented in the upper tail of flood peaks from USGS stream gaging stations in the central Appalachians, but they dominate the envelope curve of flood peaks ([*Miller*, 1990], [*Smith et al.*, 1996], [*Hicks et al.*, 2005] and [*Smith et al.*, 2011b]), based on indirect discharge estimates. Warm season thunderstorms account for a handfull of the flood events with upper tail ratio greater than 5 in the eastern US (Figure 2), including the Smethport flood of 18-19 July 1942, the Rapidan flood of 27 June 1995 and the northern Ohio flood of 4 July 1969 (see [*Smith et al.*, 2011b]). The Smethport storms of 18-19 July 1942 in northwestern Pennsylvania ([*Eisenlohr*, 1952] and [*Smith et al.*, 2011b]) produced the world record rainfall of 780 mm in 4.75 hours and flood peaks (based on indirect discharge measurements) that control the eastern US envelope curve at basin scales less than 100 km^2. They were exceptionally intense thunderstorms; in transmitting rainfall analyses for the [*Eisenlohr*, 1952] report, J. E. Stewart notes that "everyone contacted emphasizes that this was by far the worst electrical storm they ever witnessed, not only for length of time involved, but particularly for intensity at the time any one of the thunderstorms were passing" (J. E. Stewart, letter to Merill Bernard, Supervising Hydrologist, U. S. Weather Bureau, 18 September 1942).

4 The Strangest Floods: The Blue Mountains of Eastern Oregon

In this section we examine the hydrology, hydrometeorology and hydroclimatology of extreme floods in and adjacent to the Blue Mountains of eastern Oregon. The region, which is located along the southern boundary of the Columbia Plateau, is the setting for the largest upper tail ratio flood in the USGS stream gaging record (excluding the Teton dam failure flood peak) and two "miscellaneous" flood peaks which control the US and world envelope curve at 10 - 40 km^2 scale (see Figure 12 for locations). The Columbia Plateau is an apt setting for strange floods. The largest floods on Earth, the Lake Missoula floods, shaped the landscape of eastern Washington and Oregon 15,000 - 13,000 years ago (see, for example, [*O'Connor and Baker*, 1992]). In a paper that examines the properties of strange floods, this section is designed to highlight key features of the strangest.

4.1 Heppner Flood

What is known about rainfall and streamflow from the Heppner flood comes largely from USGS hydrologist John Whistler ([*Whistler*, 1903]; a synopsis of Whistler's report appears in [*Murphy*, 1904]). Whistler introduced the event by noting that "On June 14 there occurred at Heppner, Oregon, one of those storms so peculiar to arid regions and generally referred to as 'cloud-bursts'. A knowledge of the amount of precipitation and

673 the area covered, seemed to be of special interest in the consideration of irrigation projects
674 ... I have made an investigation to determine more or less approximately the facts as to
675 run-off and area of storm, ... together with other facts which I have thought may be of
676 general interest".

677 The 1903 flood in Willow Creek reached Heppner (see Figure 13) at 5:30 pm lo-
678 cal time and resulted in more than 250 fatalities in less than 1 hour ([*Whistler*, 1903]).
679 The time of "first water" and flood crest were almost simultaneous for Willow Creek at
680 Heppner. The floodwaters came principally from Balm Fork, a 68 km^2 tributary of Wil-
681 low Creek. The 14 June 1903 flood peak in Balm Fork (Figure 1) was estimated to be
682 36,000 cfs (1019 $m^3\ s^{-1}$), 198 times larger than the sample 10-year flood magnitude (Ta-
683 ble 1). The unit discharge for the Balm Fork peak is 15 $m^3\ s^{-1}\ km^{-2}$. The estimated
684 flood peak in Willow Creek at Heppner (Figure 13), which has a drainage area of 251
685 km^2, was also 36,000 cfs (1019 $m^3\ s^{-1}$), 78 times larger than the sample 10-year flood
686 magnitude. The peak discharge values for Balm Fork and locations downstream on Wil-
687 low Creek were based on slope-area computations ([*Whistler*, 1903]; see [*Levish and Os-*
688 *tenaa*, 1996] and [*Byrd*, 2009] for additional discussion and photographs).

689 The Heppner flood was the product of a hailstorm that moved down the valley of
690 Balm Fork. Whistler summarized the Heppner storm as follows: "residents of the val-
691 ley and other eyewitnesses agree that the storm moved down the valley of Balm Fork ...
692 and had entirely passed away in less than a half-hour. Accounts agree that it extended
693 from the foot of the Blue Mountains, 8 to 10 miles distant, to and a short distance across
694 Willow Creek valley, covering a strip, variously said to be from 2 to 4 miles wide." Whistler
695 concluded that rainfall at any location was no more than 30 minutes in duration.

696 We use the KINEROS-2 model, which has been applied to flood modeling for arid/semi-
697 arid watersheds (see [*Morin et al.*, 2006] and [*Goodrich et al.*, 2011] for additional de-
698 tails), to examine the dependence of flood peaks on structure, motion and rainfall rates
699 from the Heppner storm. We use a 30 meter USGS DEM for implementing KINEROS-
700 2 for Balm Fork (see Figure 13), with the 68 km^2 watershed partitioned into 117 over-
701 land flow planes and 77 channel elements. Whistler describes the hillslopes of Balm Fork
702 as "rocky and barren hills which carry off water very rapidly". Based on Whistler's anal-
703 yses, we assume low values of saturated hydraulic conductivity (0.001 $mm\ h^{-1}$) and as-
704 sume channel roughness of 0.04, in line with Whistler's assumptions for peak discharge
705 computations. Channel geometry is also based on Whistler's analyses. We use the AGWA
706 algorithms ([*Goodrich et al.*, 2011]) with conventional GIS data layers to specify model
707 parameters that are not fixed by observations from the 1903 flood (see [*Kampf and Burges*,
708 2007] and [*Goodrich et al.*, 2011] for discussion of errors inherent in modeling analyses
709 with KINEROS and similar hydrologic models).

710 As noted above, Whistler's report provides an idealized depiction of storm size and
711 motion for the June 1903 Heppner storm. From these, we develop rainfall rate fields for
712 the June 1903 storm, representing the storm as a circular region of uniform rainfall rate
713 and radius 5 km, which moves down the Balm Fork watershed, i.e. from southeast to north-
714 west, at a speed resulting in 30 minute rainfall durations for locations in Balm Fork. A
715 model simulation based on rainfall fields consistent with Whistler's characterizations of
716 storm structure and motion and with a rainfall rate of 135 $mm\ h^{-1}$ reproduces the peak
717 discharge and timing of peak response (Figure 14). In Figure 14, we also show sensitiv-
718 ity of model analyses to assumptions on rainfall rate and storm motion. Rainfall rates
719 at the 30-minute time scale have return intervals that are large for eastern Oregon (the
720 100-year 30-minute rainfall rate for Heppner is approximately 40 $mm\ h^{-1}$, based on Ore-
721 gon Department of Transportation analyses), but are not uncommon for hailstorms pro-
722 ducing extreme flooding in small watersheds ([*Doswell et al.*, 1996] and [*Smith et al.*, 2001]).
723 As will be shown below, 30-minute rainfall rates needed to produce peak flood response
724 in Balm Fork are smaller than those required to produce estimated flood peaks for the
725 Meyers Canyon and Lane Canyon events, based on similar modeling analyses.

Model simulations suggest that storm motion played a siginficant role in determining flood peak magnitude in Balm Fork. For a circular storm of radius 5 km (approximately 50 km^2 in area) moving down Balm Fork with a rain rate of 100 $mm\ h^{-1}$, the peak discharge at the outlet is 931 $m^3\ s^{-1}$ (Figure 14). For a 5 km radius storm moving up the basin at the same speed the peak discharge is 485 $m^3\ s^{-1}$. In both cases the storm total rainfall and distribution of instantaneous rain rates over the watershed are the same. We note below that for typical thunderstorm environments in eastern Oregon, "steering winds" have a southerly component. The combination of a preferred pattern of storm motion and sensitivity of extreme flood peaks to storm motion relative to the drainage network, suggests that topographic heterogeneity can play an important role in determining regional variation of the upper tail of flood distributions. Not only are the Blue Mountains a distinctive setting for extreme floods, but locations like the Willow Creek watershed will have enhanced potental for flooding due to properties of storm motion relative to the drainage network structure (see [*Soderholm et al.*, 2014] for related analyses of the Black Hills region of South Dakota). "Hotspots" of catastrophic flooding in mountainous terrain have been examined in geomorphic studies of extreme floods and debris flows ([*Eaton et al.*, 2003]; see also [*Hack and Goodlett*, 1960]) and in paleoflood reconstructions ([*Harden et al.*, 2011]).

The June 1903 storm is not entirely unprecedented for the Willow Creek watershed. Major flooding occurred on 25 May 1971 from a storm in the Shobe Creek tributary to Willow Creek. Shobe Creek enters Willow Creek near Heppner downstream of the confluence of Balm Fork (Figure 13). A USGS indirect discharge measurement for the 25 May 1971 flood peak in Shobe Creek puts the peak discharge at 172 $m^3\ s^{-1}$ from a drainage area of 16 km^2. Like the Heppner storm, the 1971 Shobe Creek storm was a hailstorm with large hail and heavy rain. The Heppner Gazette-Times of 27 May 1971 reported that "a sudden hard torrent of rainfall with a brief dash of lightning struck about 3:15 PM ... The alfalfa and grain crops were badly damaged by hail [accumulations of four inches] ... This was followed by a hard rain".

The 1903 Heppner storm was an unusually intense hailstorm. Whistler notes that rainfall was preceded by a "phenomenal fall of hail" and that "grim evidence of the amount of hail is that, while most of the bodies being recovered on the fifth day were already badly decomposed, one was occasionally found almost perfectly preserved in a large drift of hail" ([*Whistler*, 1903]). Eyewitness accounts of the storm suggest that it may have been a supercell thunderstorm, with one observer describing the storm as "a broad black streak almost straight north of us which seemed to be coming straight down from the sky" ([*Byrd*, 2009]). Supercells are the most intense storms in the spectrum of thunderstorms and have produced some of the largest short-term rainfall rates in the US ([*Giordano and Fritsch*, 1991], [*Doswell et al.*, 1996], [*Smith et al.*, 2001], [*Rogash and Racy*, 2002], [*Nielsen et al.*, 2015] and [*Bunkers and Doswell*, 2016]). Recent studies are shedding light on both the dynamics and climatology of supercells in mountainous terrain (see, for example, [*Bosart et al.*, 2006], [*Markowski and Dotzek*, 2011] and [*Scheffknecht et al.*, 2017]).

The NCDC Storm Events database and USGS annual peak discharge record contain no events in the Blue Mountains region that are comparable to the Heppner storm during the period of modern observing records, which we take to begin with deployment of the WSR-88D radar network in the mid-1990s. One of the closest analogs is a hailstorm which passed to the west of Heppner on 17 July 2012. Like the Heppner storm, the 17 July 2012 storm was an intense hailstorm embedded in a multi-day period of thunderstorm activity. Polarimetric radar analyses of the 17 July 2012 storm (Figures 15 and 16) illustrate structure, motion and evolution of severe hailstorms. The 17 July 2012 storm initiated along the Blue Mountains, as the Heppner storm did, and generally moved from southeast to northwest as the Heppner storm did, passing along the western margin of Balm Fork and producing a (small) annual peak discharge at the Balm Fork stream gaging station. The NCDC Storm Events database includes reports of heavy rain, more than

779 1 inch in 1 hour as the storm passed Shobe Canyon (approximately 2335 UTC; see Fig-
780 ure 15), and multiple reports of 1 - 2 inch hail after 2315 UTC.

781 The 17 July 2012 storm was anything but uniform in its structure, motion and evo-
782 lution (Figures 15 and 16; storm tracking analyses are based on the TITAN algorithms
783 of [*Dixon and Wiener*, 1993]). The peak value of reflectivity at 2335 UTC was 73 dBZ
784 (Figure 16) and the paired value of differential reflectivity was close to 0, indicative of
785 large hail. The storm decayed rapidly around 0015 UTC, as it moved north of Heppner
786 towards the Columbia River. Merger of two cells around 23:20 UTC resulted in an in-
787 crease in storm size and the spike in maximum reflectivity around 23:30 UTC (Figure
788 16). Storm area (determined by 45 dBZ threshold) ranged from 20 to 80 km^2, with a
789 sharp increase in size after merger of the two storm elements. Storm motion was deter-
790 mined by storm propagation and advection by steering winds, with storm speed fluctu-
791 ating around 30 $km\ h^{-1}$. Polarimetric radar measurements of specific differential phase
792 shift (Figure 15; see [*Romine et al.*, 2008] and [*Kumjian*, 2013]), which is approximately
793 linearly related to rainfall rate and immune to the effects of hail, illustrate the rapidly
794 varying temporal and spatial structure of rainfall. Polarimetric measurements indicate
795 short periods and small areas with rainfall rates of approximately 100 $mm\ h^{-1}$, corre-
796 sponding to specific differential phases shift values larger than 3 degrees km^{-1} (Figure
797 15). Analyses of the 17 July 2012 storm based on modern observing capabilities point
798 to the large uncertainties in characterizing rainfall from the Heppner storm (as described
799 above) and from the Meyers Canyon and Lane Canyon storms (as detailed below).

800 **4.2 Meyers Canyon and Lane Canyon Floods**

801 The Meyers Canyon and Lane Canyon floods in eastern Oregon (see Figure 12 for
802 locations) control the US and world envelope curves at basin scales ranging from 10 to
803 40 km^2. Both the Meyers Canyon and Lane Canyon flood peaks resulted in larger unit
804 discharge peaks than the Balm Fork flood of 1903. Perhaps more than any other floods
805 in the USGS record, they highlight the uncertainties in measurement of extreme flood
806 magnitudes (see [*Costa and Jarrett*, 2008] for discussion of the accuracy of both the Mey-
807 ers Canyon and Lane Canyon flood peak measurements). Unlike Balm Fork, no system-
808 atic stream gaging has been established by the USGS at either the Meyers Canyon or
809 Lane Canyon sites.

810 The estimated peak discharge for the Meyers Canyon flood of 13 July 1956 ([*Hen-*
811 *dricks*, 1964]; see Figure 12 for the location) is 1543 $m^3\ s^{-1}$ (54500 cfs) from a drainage
812 area of 32.9 km^2 (12.7 mi^2), resulting in a unit discharge of 46.9 $m^3\ s^{-1}\ km^{-2}$ (4290
813 cfs per square mile). A rough estimate of the upper tail ratio for the Meyers Canyon flood
814 peak is 600, which we obtain using the regression equations developed by *Cooper* [2006]
815 to estimate the 10-year flood. The accuracy of the flood peak estimate has been, and con-
816 tinues to be, a topic of debate (see, for example, [*Levish and Ostenaa*, 1996]). Follow-
817 ing site investigation and review of previous analyses, *Costa and Jarrett* [2008] concluded
818 that the "54500 cfs peak should remain as published, but the rating should be downgraded
819 from fair to estimate", which implies that the uncertainty in the estimate is larger than
820 25% and potentially as large as 50%. Even if peak discharge were at the lower end of the
821 uncertainty range, it would rank with the June 1903 Balm Fork peak as one of the most
822 extreme flood measurements in the US.

823 An observer reported that rainfall in Meyers Canyon lasted for approximately 2
824 hours (4:30 pm local time to 6:30 pm local time), but a large part of the precipitation
825 occurred in approximatly 30 minutes with rainfall in "cloudburst proportions" ([*Hen-*
826 *dricks*, 1964]). There are no reports of lightning or hail for the storm, but the area is re-
827 mote and there was only a single report from the area of heavy rainfall in Meyers Canyon.
828 Downscaling simulations with WRF (as detailed below) indicate that the storm environ-

ment, like the Heppner and Lane Canyon settings, would support the development of thunderstorms.

KINEROS-2 simulations were carried out for the Meyers Canyon flood, emulating the procedures used for the Heppner flood. A rainfall rate of 250 $mm\ h^{-1}$ from a storm moving down the Meyers Canyon drainage network at 32 $km\ h^{-1}$ produces a 30-minute period of rainfall for locations in the watershed and a peak discharge close to 1543 $m^3\ s^{-1}$ at the basin outlet. The peak discharge that results from the same scenario, except for motion up the watershed, is 637 $m^3\ s^{-1}$ at the basin outlet.

The Lane Canyon flood on 26 July 1965 (see the summary by D. D. Harris in [*Rostvedt*, 1970]) produced a flood peak of 807 $m^3\ s^{-1}$ (28500 cfs) from a drainage area of 13.1 km^2 (5.04 mi^2), resulting in a unit discharge of 61.6 $m^3\ s^{-1}\ km^{-2}$ (5650 cfs per square mile). Using the regression equations of *Cooper* [2006] to estimate the 10-year flood magnitude in Lane Canyon, yields an upper tail ratio of approximately 800 for the 1965 flood. Accuracy of the estimated peak discharge for the Lane Canyon flood, like the Meyers Canyon peak, has sparked controversy. *Costa and Jarrett* [2008] concluded that "the peak discharge for the flood in Lane Canyon was likely 28500 cfs, but the rating should be changed from fair to poor". Also like the Meyers Canyon flood, the Lane Canyon flood peak would be one of the most extreme in the US even if peak discharge were at the low end of discharge values associated with measurement uncertainty.

The Lane Canyon flood was caused by a "severe cloudburst", with observers reporting that most of the rain fell in 30 minutes ([*Rostvedt*, 1970]). Observers also reported that large accumulations of hail accompanied the storm. The spatial extent of erosional features indicated that the entire watershed produced heavy runoff. KINEROS-2 simulations for a storm moving down the watershed with a rain rate of 360 $mm\ h^{-1}$ produce a peak discharge close to the observed maximum of 807 $m^3\ s^{-1}$. The same storm moving up the watershed produces a peak discharge of 330 $m^3\ s^{-1}$.

Using comparable model implementations, we conclude that the rainfall rates needed to produce the estimated peak discharge values for the 1956 Meyers Canyon and 1965 Lane Canyon flood peaks are larger than those required to produce the Balm Fork peak. The rainfall rates are not unprecedented for thunderstorms in the US ([*Engelbrecht and Brancato*, 1959] and [*Smith et al.*, 2011b]), but they are exceptionally rare. These analyses highlight the large uncertainties in discharge estimates for the Meyers Canyon and Lane Canyon flood peaks.

4.3 Hydrometeorology of Blue Mountain Thunderstorms

In this section we examine the hydrometeorology of extreme rainfall in the Blue Mountains, principally through analyses of WRF simulations of the June 1903 Heppner storm, the July 1956 Meyers Canyon storm and the July 1965 Lane Canyon storm. We first note that the climatology of thunderstorms, as represented by cloud-to-ground (CG) lightning observations from National Lightning Detection Network observations over eastern Oregon and Washington state, exhibits pronounced spatial heterogeneity, with a peak occurrence over the Blue Mountains (Figure 12). Mean annual CG flash density ranges from 0.8 strikes km^{-2} to 1.2 strikes km^{-2} along the Blue Mountains. Mean flash density decreases sharply in the Columbia Plateau region to the northwest, with a mean flash density of 0.5 km^{-2} in Heppner and values less than 0.2 km^{-2} over the Columbia River. For the WRF simulations of thunderstorms over the Blue Mountains region, we use three one-way nested domains of 27, 9 and 3 km resolution. Model physics options include: 1) Noah land surface model, 2) WSM6 microphysics scheme, 3) Yonsei planetary boundary layer scheme, 4) Dudhia shortwave radiation scheme, and 5) the RRTM (Rapid Radiative Transfer Model) scheme for longwave radiation (see [*Smith and Baeck*, 2015] for similar implementations). For the 1903 and 1965 simulations, 20th Century Reanalysis

fields were used for initial and boundary conditions. For the 1956 Meyers Canyon storm NCEP-NCAR reanalysis fields were used for initial and boundary conditions.

The simulation for the June 1903 storm captures the occurrence of heavy rainfall along the Blue Mountains (Figure 17). Peak rainfall accumulations of 80-100 mm are located in the higher elevations of the Blue Mountains east of Heppner, with rainfall near Heppner exceeding 50 mm. Broad features of the simulated rainfall are consistent with the available evidence on rainfall distribution. The Daily Weather Map for 14 June 1903 shows an area of rainfall extending through eastern Oregon from southern Idaho into Washington state. Simulated rainfall is organized into small convective elements (median size of 50 km^2 based on TITAN storm tracking analyses of simulated reflectivity fields) that have a south-to-north component of storm motion. The synoptic scale environment of the Heppner storm was characterized by an upper level trough in the eastern Pacific off the coast of Canada and a high pressure ridge extending from Mexico through the Rocky Mountain region into Canada.

Water vapor for the Heppner storm was exceptionally large for the Blue Mountains region. The simulated precipitable water field at 1800 UTC on 14 June 1903 shows a large area of extreme precipitable water values centered on the Columbia Plateau (Figure 18). Precipitable water values approached 40 mm, which is close to the upper bound of precipitable water observations from the GPS precipitable water station near Heppner, Oregon (Figure 19). Paired with elevated precipitable water, the WRF simulation indicates that the convective potential was large over a significant portion of the Columbia basin margin, with peak values of Convective Available Potential Energy (CAPE) approaching 1000 $J\ kg^{-1}$ adjacent to the highest elevation region of the Blue Mountains in eastern Oregon (figure not shown). The 0.9 quantile of CAPE for June at the nearest radiosonde site in Spokane, Washington is less than 300 $J\ kg^{-1}$ and the record CAPE for June is less than 2000 $J\ kg^{-1}$ based on observations extending back to 1948. Like the Meyers Canyon and Lane Canyon events, the Heppner storm occurred during a multi-day period of elevated humidity and convective potential. A tornado was reported the day after the Heppner storm (15 June 1903) approximately 57 miles northeast of Heppner ([*Byrd*, 2009]).

A central issue surrounding record flooding in the arid region of eastern Oregon is the source of water vapor that fueled the storms. Back-trajectory analyses using the WRF simulation and the LAGRANTO algorithm (Figure 20; see [*Sprenger and Wernli*, 2015] for details on back-trajectory computations) illustrate a complex pattern of water vapor transport. Back-trajectory analyses ending at 1800 UTC on June 14 and at 5 km elevation show parcel transport from southwest to northeast from the Pacific to eastern Oregon. Similar analyses terminated at 2 km elevation over eastern Oregon show diverse flowpaths to the Blue Mountains flood region, including transport from Idaho and Montana. Unlike major flood events in the Pacific Northwest from cold-season extratropical systems, the Heppner storm was not the product of a well-defined "river" of atmospheric water vapor transport (*Colle and Mass* [2000] and [*Neiman et al.*, 2011]).

Simulated rainfall for the July 1965 Lane Canyon storm includes a local maximum (70 mm) in rainfall on the northern margin of the Blue Mountains (Figure 21). The US Weather Bureau surface map shows an area of rainfall that matches the pattern of rain extending from northeastern Oregon into northwestern Montana. Like the Heppner storm, the Lane Canyon storm environment was characterized by extreme values of precipitable water (Figure 22) and elevated CAPE (Figure 23) along the margins of the Columbia Plateau. The synoptic environment for the Lane Canyon storm was similar to the Heppner storm, with comparable trough/ridge positions.

The WRF simulation of the 1956 Meyers Canyon storm (Figure 21) captures key elements of rainfall for the day, as represented by the US Weather Bureau surface map, which indicated areas of rainfall extending along the California - Oregon border north-

eastward into Oregon. A local maximum in the WRF simulation is located southwest of Meyers Canyon. The largest observed rainfall for the day was located along the Montana - Canada border in the area of the local maximum from the WRF simulation. Precipitable water and CAPE fields for the Meyers Canyon storm are similar to the Heppner and Lane Canyon storm environments (Figures 22 and 23). The synoptic environment for the Meyers Canyon storm was characterized by an upper level trough that extended along the northern California coast into Oregon and Washigton.

The synoptic environment of the 17 July 2012 storm, the modern analog of the Heppner storm, was similar to the Meyers Canyon storm, with an upper level trough that extended from north to south along the Pacific Northwest during the period from 16 - 18 July 2012. Like the Heppner, Lane Canyon and Meyers Canyon environments, elevated precipitable water characterized a 4-day period of thunderstorms in eastern Oregon and Washington, with peak values of precipitable water at Heppner reaching 35 mm. The period from 15-18 July 2012 contains 4 of the top 10 thunderstorm days in the Blue Mountain region since 1992, based on NLDN CG flash density observations. In addition to copious lightning, storm reports for the period include large hail, funnel clouds, heavy rain, flash floods, debris flows and landslides for eastern Oregon and Washington (NCDC Storm Reports).

Downscaling simulations with WRF provide insights to common elements of storm environment for the record floods of Eastern Oregon, reducing some of the strangeness of these events. The Heppner, Meyers Canyon and Lane Canyon storms exhibited similar structure of steering level winds, precipitable water and CAPE fields, which play an important role in determining the potential for extreme rainfall rates (see [*Nielsen et al.*, 2015], [*Bunkers and Doswell*, 2016] and [*Nielsen and Schumacher*, 2018] for additional discussion). Model simulations also suggest that storm size and motion were similar for the three events. Storm size and motion dictate that extreme floods in the Blue Mountains region will concentrate at basin scales smaller than 100 km^2. The Heppner, Meyers Canyon and Lane Canyon floods are rare, but the environmental conditions for multi-day thunderstorm outbreaks in the Blue Mountains region are recurring phenomena, as illustrated by the July 2012 storm period. The WRF simulations suggest that some aspects of these major thunderstorm outbreaks are predictable, based on information contained in conventional reanalysis fields. A mystery that remains is how some storms that form during major thunderstorm outbreaks can produce rainfall rates that result in Heppner-like flood peaks.

5 Summary and Conclusions

We examine the upper tail of flood peaks in the US, principally through analyses of the upper tail ratio for annual flood peak observations from more than 8000 USGS stream gaging stations. Our major findings are the following:

- With a few caveats, the 14 June 1903 Heppner, Oregon flood is the strangest flood in the systematic stream gaging record of the US. The upper tail ratio for Balm Fork at a drainage area of 68 km^2 is 200, implying that the June 1903 flood peak was 200 times larger than the 10-year flood for the watershed. The 1903 Heppner flood is paired with two other extreme floods in the Blue Mountain region of eastern Oregon. Based on indirect discharge estimates in the "miscellaneous" flood record of the USGS, the July 1956 Meyers Canyon flood and the July 1965 Lane Canyon flood control the US and world flood peak envelope curves at basin scales between 10 and 40 km^2 (see [*Costa*, 1987a] and [*Herschy*, 2003]).
- A fundamental distinction between record floods in the US and the broader population of annual flood peaks is in their seasonal distribution. Annual flood peaks in the US have winter/spring and warm season maxima in the seasonal rate of occurrence. For record floods, there is a striking shift towards warm season flood-

ing and an increase in occurrence probability during the tropical cyclone season. This feature becomes more pronounced as we focus on floods with larger upper tail ratio. These analyses point to warm season thunderstorms and tropical cyclones as major flood agents for record floods, in contrast to the broader population of annual flood peaks, which are dominated by snowmelt and winter/spring storm systems.

- The strangest floods in the US, i.e. those producing the largest upper tail ratios (Table 1), are heavily weighted toward intense thunderstorms during the warm season. Furthermore, the most intense forms of convection, hailstorms and supercell thunderstorms, are prominent agents of strange floods. These include some of the most deadly and devastating floods in US history, notably the June 1903 Heppner flood, the May 1935 Republican River flood, the April 1934 Washita flood, the June 1965 Front Range floods and the June 1972 Rapid City flood. Severe thunderstorms in and adjacent to mountainous terrain are prominent in the largest upper tail ratio floods (Table 1). Strange floods from intense convection are not restricted to the smallest basin scales; the May 1935 Republican River flooding extended to basin scales exceeding 30,000 km^2. The largest flood peak in the USGS stream gaging record resulted from the 1927 flooding in the Lower Mississippi River at a basin scale of 1,000,000 km^2 and was the product of multiple episodes of mesoscale convective systems, which produced severe thunderstorms and tornadoes ([*Smith and Baeck*, 2015]).
- Strange floods can result from flood agents that are uncommon for the watershed or they can result from extremes of typical flood agents. The eastern Oregon floods, which are produced by severe thunderstorms in a region dominated by snowmelt are quintessential examples of the former. The Front Range and High Plains floods in Table 1 (May 1935 Republican, April 1934 Washita, June 1965 Front Range, June 1972 Rapid City and 28 May 1953 Prairie Dog Creek), all the product of mesoscale convective sytems during late spring and early summer, are examples of the latter. We introduce the record flood peak seasonality as a tool for examining anomalies in the seasonal occurrence of record flood peaks.
- We show that under the assumption that the annual flood peak distribution for a site is GEV with location μ, scale σ and shape ξ, the upper tail ratio has a GEV distribution that depends only on the shape parameter ξ and sample size. Estimates of the location and scale parameters for stationary USGS flood records are linearly related and they are linearly related to drainage basin area. Estimated shape parameters show no dependence on basin area or any other basin property. The distribution of upper tail ratio does not depend on drainage area, provided that sample size is accounted for, supporting a scaling theory for the upper tail ratio that is dependent only on the shape parameter of the parent GEV distribution and sample size. Upper tail ratio analyses for stationary USGS stations are consistent with an unbounded, thick-tailed GEV world. Surprises, in the form of unprecedented flood peaks, are a prominent feature of a thick-tailed GEV world. Error bounds in estimates of extreme flood quantiles will increase with increasing values of the GEV shape parameter.
- We examine the geography of record floods by conditioning the upper tail ratio on basin properties. Arid / semi-arid watersheds have elevated values of upper tail ratio. The distribution of upper tail ratio values decreases with impervious fraction. Urban watersheds have a high frequency of large unit discharge flood peaks, but the maximum flow can be limited by capacity constraints in urban drainage networks. Jarrett's hypothesis that large floods do not occur above 2300 meters is an important element of the hydroclimatology of upper tail ratio values. There is no evidence that forest cover affects the distribution of upper tail ratios in watersheds of the US.
- Nonstationarities are a prominent feature of some stream gaging stations with large upper tail ratio values. The Republican River flood of May 1935 provides a pro-

totypical example. Devastating flooding lead to river regulation with flood control dams and levees and a coupled program of stream gaging. This pattern and closely related variants of this pattern characterize the flood series for a number of the USGS stream gaging stations in Table 1.

- Large upper tail ratio flood events in the eastern US are dominated by tropical cyclones. More than 2/3 of the record flood peaks with upper tail ratios greater than 5 east of the Mississippi River were produced by tropical cyclones. Hurricanes Diane (1955), the New England Hurricane of 1938, Hurricane Floyd (1999), Hurricane Alberto (1994) and Hurricane Agnes (1972) account for 40% of the upper tail ratios larger than 5 in the eastern US.
- Record floods in June 1903, July 1956 and July 1965 in the Blue Mountains of Oregon reflect two of the most important aspects of strange floods in the US: they are located in and adjacent to mountainous terrain and intense thunderstorms are the principal agents of extreme rainfall. The storm environments for the three floods include similar patterns of precipitable water, convective available potential energy (CAPE) and steering winds around the Blue Mountains, with near-record values of precipitable water for each event. Transport of water vapor to the Blue Hills region occurs over 3-5 days and can involve diverse and complex paths to the region of heavy rainfall. Each of the flood events was embedded in a mulit-day period of thunderstorms in eastern Oregon and Washington. The 15-18 July 2012 thunderstorm outbreak provides a modern analog for the Blue Mountains flood events.
- Hydrologic modeling analyses show that motion down the drainage network of Balm Fork, which was documented by *Whistler* [1903] based on observer accounts, results in much larger peak discharge at Heppner than for a similar storm that moves up the drainage network. Modeling analyses for Meyers Canyon and Lane Canyon result in extreme rainfall rates, highlighting the uncertainty of the peak discharge estimates for these sites. The magnitudes of peak rainfall rates are unknown for the three storms and we have a poor understanding of the physical processes that in rare circumstances produce exceptionally large rainfall rates.

Eastern Oregon is an interesting setting for examining the regional variation of extreme floods, because it includes the location of the largest upper tail ratio flood in the US (excluding the flood peak resulting from the Teton Dam failure) and the locations of two "miscellaneous" USGS flood peak measurements that control the US and world flood peak envelope curves at basin scales ranging from 10 to 40 km^2. One reading of the hydrometeorological and hydrologic analyses of the Heppner, Meyers Canyon, Lane Canyon and July 2012 events is that spatial heterogeneity of upper tail flood properties is extremely large and that each watershed will have unique flood properties. This point of view is largely driven by hydrometeorological properties of storms that control extreme floods, including the environmental conditions that create the distinctive patterns of precipitable water, CAPE and steering winds for the episodic thunderstorm outbreaks in the region. Extreme floods in the Blue Mountains reflect an interplay between storm motion and drainage network structure, providing strong linkage between hydrometeorological and hydrologic controls of upper tail flood properties.

Regionalization is one of the core elements of many techniques used for rainfall and flood frequency analysis, including procedures used for assessing Probable Maximum Precipitation (PMP) and Probable Maximum Flood. A conclusion that extreme floods in watersheds of the Blue Mountains exhibit large spatial heterogeneity does not alter the need for, and utility of, regionalization procedures. But, the Blue Mountain analyses do suggest that major advances in regional assessments of upper tail flood properties require advances in understanding and modeling hydrometeorological processes that control extreme rainfall from thunderstorms. The early ideas underlying hydrometeorological procedures for PMP analysis were driven by the goal of developing modeling capabilities that

would be coupled with appropriate observations to provide local assessments of maximum precipitation ([*Myers*, 1967]). A solid foundation for regionalization will require advances in hydrometeorology along the lines envisioned many decades ago.

The themes that emerge from examination of regional variation of extreme floods in the Blue Mountains region of Oregon are likely prominent features of many (if not most) settings in the US. Landfalling tropical cyclones in Texas produce extreme floods which are strongly dependent on amplification of rainfall at the land - water boundary (as was the case for Hurricane Harvey in August of 2017) and at inland locations where complex terrain has contributed to record rainfall accumulations for the US. Similar features characterize tropical cyclone flood hazards in coastal and inland regions along the remainder of the Gulf Coast and the east coast of the US. The mountainous regions with large upper tail ratio flood peaks (Section 3.b) are also notable settings in which spatial variation of upper tail flood properties has received considerable attention; these include the Front Range of the Rocky Mountains, the Black Hills, the North American Monsoon region of the Southwestern US, central Appalachians and the San Gabriel Mountains of California. Geomorphic and paleohydrologic studies have provided insights to "hotspots" of extreme flood occurrences and are key tools for advancing our understanding of the geography of upper tail flood properties.

One of the biggest challenges in developing a deeper understanding of strange floods centers on the physical processes controlling rainfall rates in thunderstorms. We do not know how large the rainfall rates were from the Heppner, Meyers Canyon and Lane Canyon storms. The rain rates inferred from flood peak values and hydrologic modeling for the Meyers Canyon and Lane Canyon storms, which are roughly in the range from 250 - 400 $mm\ h^{-1}$ certainly seem too large, but it is difficult to constrain rainfall rates from thunderstorms based on rain gage or even conventional radar reflectivity measurements for any setting, especially sparsely monitored locations like those in eastern Oregon. There have been larger rainfall rates documented for thunderstorms in the US. The 18-19 July 1942 Smethport, Pennsylvania storm not only produced the world record rainfall rate measurement of 164 $mm\ h^{-1}$ for 4.75 hours (780 mm accumulation), but also rain rates at 5-10 minute time scale of 1000 $mm\ h^{-1}$ ([*Eisenlohr*, 1952]). How does that happen? Polarimetric radar measurements (Figures 15) hold significant potential for examining physical processes that control rainfall rate in thunderstorms ([*Kumjian*, 2013]), because they can circumvent problems associated with hail and provide observational constraints on the interplay between storm dynamics and microphysics. Considerable progress has been made in understanding the dynamics of hailstorms, especially tornadic supercell storms, but relatively little attention has been focused on understanding the physical processes that control extreme rainfall rates, despite the fact that flash flooding is a comparable hazard to tornadoes in terms of loss of life. The quasi-idealized simulations of rainfall from supercell thunderstorms in *Nielsen and Schumacher* [2018] provide an important avenue for advancing the predictive understanding of extreme rainfall rates.

Our analyses of USGS flood records suggest that the nature of flood peak distributions in the conterminous US is unbounded and thick-tailed; we certainly find little support for Horton's argument that rainfall and flood peaks are bounded. There is a substantial population of gaged watersheds for which the upper tail ratio is greater than 10 (Figure 2) and the Heppner flood provides evidence of upper tail ratios greater than 100. Comparisons between distributional properties of upper tail ratio values in the US with other settings around the world will provide additional insights to the nature of the upper tail of flood peaks.

1196 **Table 1.** US flood events with Upper Tail Ratio greater than 20.

Name	Date / USGS ID	Upper Tail Ratio	Area km^2	S_i
1. Teton Dam	5 June 1976			
Teton River	13055000	**319**	2295	NA
2. Heppner	14 June 1903			
Balm Fork	14034480	**200**	68	1.288
Willow Creek	14034500	**80**	252	1.560
3. Republican	31 May 1935			
Republican R.	06837000	**71**	31752	3.002
Republican R.	06843500	**35**	37413	3.369
Republican R.	06828500	**32**	21007	2.962
Red Willow Ck.	06837500	**25**	1808	2.202
4. Washita	3 April 1934			
Washita R.	07324200	**57**	3539	0.772
5. Front Range	16-17 June 1965			
Plum C.	06709500	**55**	783	1.769
Jimmy Camp Ck.	07105900	**34**	170	2.193
Cimmaron R.	07207000	**27**	708	1.976
Rayado Ck.	07208500	**21**	159	1.470
6. Rapid City	9 June 1972			
Rapid Ck..	06412500	**49**	973	4.095
Rapid Ck.	06414000	**34**	1075	3.673
Boxelder Ck.	06422500	**32**	244	3.737
Spring Ck.	06408500	**26**	531	3.949
Battle Ck.	06404000	**21**	152	4.961
7. Prairie Dog	28 May 1953			
Prairie Dog Ck.	06847900	**36**	1536	1.998
8. Sybille Ck.	20 Aug 1990			
Sybille Ck.	06664400	**35**	505	1.433
9. San Felipe CA	21 Feb 1980			
San Felipe Ck.	10255700	**27**	255	1.131
10. Ashland, OR	15 Jan 1974			
EF Ashland Ck.	14353500	**25**	20.9	0.860
WF Ashland Ck.	14353000	**21**	25.5	0.856
11. San Antonio CA	2 March 1938			
San Antonio Ck.	11073000	**24**	43	1.653
12. Sand Creek CO	19 July 1977			
Sand Ck.	06659580	**22**	79.1	0.919
13. Costilla NM	22 July 1954			
Costilla Ck.	08252500	**22**	65.6	2.082
14. Medio TX	22 Sept 1967			
Medio Ck.	08189300	**21**	527	1.155
15. Teton River	8 June 1964			
Teton River	06102500	**21**	287	11.20
16. San Emigdio CA	6 Aug 1961			
San Emigdio Ck.	11195500	**20**	127	0.659

A: Appendix

The GEV family of distributions satisfies the max-stable property ([*Leadbetter et al.*, 1982]) implying that the maximum flood $X_{(n)}$ has a GEV distribution that is related to the distribution of the annual peaks X_i, through a transformation of the scale a_n and location b_n:

$$X_{(n)} \sim a_n X_i + b_n \tag{A.1}$$

For $\xi \neq 0$, $a_n = n^{\xi}$ and $b_n = (\mu - \frac{\sigma}{\xi})(1 - n^{\xi})$ and for the Gumbel case the sequences simplify to $a_n = 1$ and $b_n = \mu + \sigma\, ln(n)$.

For $X_i \sim GEV(\mu, \sigma, \xi)$,

$$a X_i + b \sim GEV(a\,\mu + b,\ a\,\sigma,\ \xi) \tag{A.2}$$

The distribution of the record flood is thus:

$$X_{(n)} \sim GEV(\mu - \frac{\sigma}{\xi}(1 - n^{\xi}),\ n^{\xi}\,\sigma,\ \xi) \tag{A.3}$$

We assume that the sample quantile function provides a good approximation to the quantile function and evaluate the ratio of the record flood $X_{(n)}$ and the 10-year flood Q(.9). The 10-year flood under the GEV assumption is:

$$Q(.9) = \mu - \frac{\sigma}{\xi}(1 - [-ln(0.9)]^{-\xi}) \tag{A.4}$$

As noted in Section 2, estimated location and scale parameters for annual peak flood records are linearly related (and they have an intercept that is approximately 0), so we can assume

$$\mu = c\,\sigma \tag{A.5}$$

Combining these results we conclude that in a GEV world the upper tail ratio depends only on the shape parameter and sample size. In particular

$$Z \sim GEV(\frac{c - g(\xi)}{c - h(\xi)},\ \frac{n^{\xi}}{c - h(\xi)},\ \xi) \tag{A.6}$$

where

$$h(\xi) = \frac{(1 - [-ln(0.9)]^{-\xi})}{\xi} \tag{A.7}$$

and

$$g(\xi) = \frac{(1 - n^{-\xi})}{\xi} \tag{A.8}$$

Acknowledgments

The authors gratefully acknowledge constructive comments Robert Mason, Jonathan Gourley and two anonymous reviewers; constructive comments were also provided by Jim O'Connor, William Asquith, Eric Gaume, Stephen Burges, Chris Milly and Gabriele Villarini on earlier versions of the paper. We also thank Dan Turner-Gillespie for the encouragement to examine the Heppner Flood and acknowledge the assistance of personnel from multiple USGS offices in sorting out details of a number of flood peak measurements. This research was supported by the National Science Foundation (NSF Grants EAR-1632048, AGS-1522492 and CBET-1444758), NASA (Grant NNX10AI46G), NOAA (Grant NA14OAR4830101) and the Moore Charitable Foundation. Annual flood peak data used in this study can

be obtained from the USGS National Water Information System. NLDN data were provided by the NASA Lightning Imaging Sensor (LIS) instrument team and the LIS data center via the Global Hydrology Resource Center (GHRC) located at the Global Hydrology and Climate Center (GHCC), Huntsville, Alabama, through a license agreement with Global Atmospherics (now Vaisala), Inc (GAI). Paul Bates was supported by a Leverhulme Research Fellowship and a Royal Society Wolfson Research Merit award.

References

Atallah, E., and L. F. Bosart (2003), The extratropical transition and precipitation distribution of Hurricane Floyd (1999), *Monthly Weather Review*, *131*, 1063–1081.

Ayalew, T. B., W. F. Krajewski, and R. Mantilla (2014), Connecting the power-law scaling structure of peak-discharges to spatially variable rainfall and catchment physical properties, *Advances in Water Resources*, *71*, 32 – 43.

Baker, V. R. (1975), Flood hazards along the Balcones Escarpment in central Texas: Alternative approaches to their recognition, mapping and management, *Tech. Rep. Bureau of Economic Geology University of Texas Circ. 75-5*, University of Texas.

Barth, N. A., G. Villarini, M. A. Nayak, and K. White (2017), Mixed populations and annual flood frequency estimates in the western United States: The role of atmospheric rivers, *Water Resources Research*, *53*(1), 257 – 269.

Billington, D. P., and D. C. Jackson (2006), *Big Dams of the New Deal Era: A Confluence of Engineering and Politics*, 369 pp., University of Oklahoma Press, Cambridge.

Blöschl, G., and M. Sivapalan (1997), Process controls on regional flood frequency: Coefficient of variation and basin scale, *Water Resources Research*, *33*(12), 2967–2980.

Boner, F. C., and F. Stermitz (1967), Floods of June 1964 in Northwestern Montana, *Water Supply Paper 1840-B*, U.S. Geological Survey, Washington D. C.

Bosart, L. F., A. Seimon, K. D. LaPenta, and M. J. Dickinson (2006), Supercell tornadogenesis over complex terrain: the Great Barrington, Massachusetts tornado on 29 May 1995, *Weather and Forecasting*, *21*, 897– 921.

Bunkers, M. J., and C. A. Doswell (2016), Comment on "Double impact: when both tornadoes and flash floods threaten the same place and time", *Weather and Forecasting*, *31*, 1715 – 1721.

Byrd, J. G. (2009), *Calamity: The Heppner Flood of 1903*, 202 pp., The University of Washington Press, Seattle, Washington.

Caracena, F., and J. M. Fritsch (1978), Focusing mechanisms in the Texas Hill Country flash floods of 1978, *Monthly Weather Review*, *111*, 2319–2332.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer.

Colle, B. A. (2003), Numerical simulations of the extratropical transition of Floyd (1999): Structural evolution and responsible mechanisms for the heavy rainfall over the Northeast U.S., *Monthly Weather Review*, *131*, 2905–2926.

Colle, B. A., and C. F. Mass (2000), The 5–9 February 1996 Flooding Event over the Pacific Northwest, *Monthly Weather Review*, *128*(3), 593 – 618.

Cooper, R. M. (2006), Estimation of Peak Discharges for Rural, Unregulated Streams in Eastern Oregon, *Open File Report SW 06-002*, State of Oregon, Salem, Oregon.

Costa, J. E. (1987a), Hydraulics and basin morphometry of the largest flash floods in the conterminous United States, *Journal of Hydrology*, *93*(3–4), 313–338.

Costa, J. E. (1987b), A comparison of the largest rainfall-runoff floods in the United States with those of the People's Republic of China and the world, *Journal of Hydrology*, *96*, 101–115.

Costa, J. E., and R. D. Jarrett (2008), *An Evaluation of Selected Extraordinary Floods in the United States Reported by the U. S. Geological Survey and Implications for Future Advancement of Flood Science*, vol. Scientific Investiations Report 2008-5164, U.S. Geological Survey.

Cotton, W. R., G. H. Bryan, and S. C. van den Heever (2010), *Storm and Cloud Dynamics*, 775 pg., 2nd edition. ed., Academic Press.

Delrieu, G., V. Ducrocq, E. Gaume, J. Nicol, O. Payrastre, E. Yates, H. A. P. E. Kirstetter and, P. A. Ayral, C. Bouvier, J. D. Creutin, M. Livet, S. Anquetin, M. Lang, L. Neppel, C. Obled, J. du Châtelet, G. Saulnier, A. Walpersdorf, and W. Wobrock (2005), The catastrophic flash-flood event of 8-9 September 2002 in the Gard Region, France: A first case study for the Cèvennes Vivarais Mediterranean Hydrometeorological Observatory, *Journal of Hydrometeorology*, *6*, 34–52.

Dettinger, M. D., F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan (2011), Atmospheric rivers, floods and the water resources of California, *Water*, *3*, 445 – 478.

DiBaldassarre, G., A. Viglione, G. Carr, J. L. Salinas, and G. Bloeshl (2013), Sociohydrology: conceptualizing human-flood interactions, *Hydrology and Earth System Sciences*, *17*, 3295 – 3303.

DiMego, G. J., and L. F. Bosart (1982), The transformation of Tropical Storm Agnes into an extratropical cyclone. Part I: The observed fields and vertical motion computations, *Monthly Weather Review*, *110*(5), 385 – 411.

Dixon, M., and G. Wiener (1993), TITAN - Thunderstorm Identification, Tracking, Analysis, and Nowcasting - A radar-based methodology, *Journal of Atmospheric and Oceanic Technology*, *10*(6), 785–797.

Doswell, C. A., H. E. Brooks, and R. A. Maddox (1996), Flash flood forecasting: An ingredients-based methodology, *Weather and Forecasting*, *11*(4), 560–581.

Douglas, E. M., and R. M. Vogel (2006), Probabilistic Behavior of Floods of Record in the United States, *Journal of Hydrologic Engineering*, *11*(5), 482 – 485.

Douglas, J. S. (1908), A California Cloudburst, *Monthly Weather Review*, *36*(9), 631–632.

Ducrocq, V., O. N. ad D. Ricard, C. Lebeaupin, and T. Thouvenin (2008), A numerical study of three catastrophic precipitation events over southern France. II: Mesoscale triggering and stationarity factors, *Quarterly Journal of the Royal Meteorological Society*, *134*, 131 – 145.

DuMouchel, W. H. (1983), Estimating the stable index α in order to measure tail thickness: A critique, *The Annals of Statistics*, *11*(4), 1019–1031.

Eaton, L. S., B. A. Morgan, R. C. Kochel, and D. H. A (2003), Quaternary deposits and landscape evolution in the central Appalachians of Virginia, *Geomorphology*, *56*, 139–154.

Eisenlohr, W. S. (1952), Floods of July 18, 1942 in north-central Pennsylvania, *U.S. Geological Water Supply Paper 1134-B*, USGS.

El Adlouni, S., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobeé (2007), Generalized maximum likelihood estimators for the nonstationary Generalized Extreme Value model, *Water Resources Research*, *43*, w03410.

Emanuel, K. (2017), Assessing the present and future probability of Hurricane Harvey's Rainfall, *Proceedings of the National Academy of Sciences*, *114*(48), 12,681 – 12,684.

Engelbrecht, H. H., and G. N. Brancato (1959), World record one-minute rainfall at Unionville, Maryland, *Monthly Weather Review*, *87*(8), 303 – 306.

England, J. F., P. Y. Julien, and M. L. Vellleux (2014), Physically-based extreme flood frequency with stochastic storm transposition and paleoflood data on large watersheds, *Journal of Hydrology*, *510*, 228 – 245.

Enzel, Y., L. L. Ely, K. House, and V. R. Baker (1993), Paleoflood evidence for a natural upper bound to flood magnitudes in the Colorado River basin, *Water*

Resources Research, *29*(7), 2287 – 2297.

Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador (2010), GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States, *Ecology*, *91*(2), 621.

Follansbee, R., and E. E. Jones (1922), Arkansas River flood of June 3 - 5, 1921, *U.S. Geological Water Supply Paper 487*, USGS.

Follansbee, R., and J. B. Spiegel (1937), Floods on the Republican and Kansas Rivers May and June 1935, *U.S. Geological Water Supply Paper 796-B*, USGS.

Fundel, F., and M. Zappa (2011), Hydrological ensemble forecasting in mesoscale catchments: sensitivity to initial conditions and value of reforecasts, *Water Resources Research*, *47*, W09,520.

Gaume, E., et al. (2009), A compilation of data on European flash floods, *Journal of Hydrology*, *367*, 70–78.

Giordano, L. A., and J. M. Fritsch (1991), Strong tornadoes and flash- flood-producing rainstorms during the warm season in the Mid-Atlantic region, *Weather and Forecasting*, *6*, 437 – 455.

Goodrich, D. C., et al. (2011), AGWA: The automated geospatial watershed assessment tool to inform rangeland management, *Rangelands*, *33*(4), 41–47.

Hack, J. T., and J. C. Goodlett (1960), Geomorphology and forest ecology of a mountain region in the central Appalachians, *U.S. Geological Prof. Paper 347*, USGS.

Halbert, K., C. C. Nguyen, O. Payrastre, and E. Gaume (2016), Reducing uncertainty in flood frequency analyses: A comparison of local and regional approaches involving information on extreme historical floods, *Journal of Hydrology*, p. DOI 10.1016/j.jhydrol.2016.01.017.

Hamada, A., Y. N. Takayabu, C. Liu, and E. J. Zipser (2015), Weak linkage between the heaviest rainfall and tallest storms, *Nature Communications*, *6*(6213), doi:10.1038/ncomms7213.

Harden, T. M., J. E. O'Connor, D. G. Driscoll, and J. F. Stamm (2011), Flood-Frequency Analyses from Paleoflood Investigations for Spring, Rapid, Boxelder, and Elk Creeks, Black Hills, Western South Dakota, *Scientific Investigations Report 2011 - 5131*, U.S. Geological Survey.

Hart, R. E., and J. L. Evans (2001), A climatology of the extratropical transition of Atlantic tropical cyclones, *Journal of Climate*, *14*, 546–564.

Helmus, J. J., and S. M. Collis (2016), The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language, *Journal of Open Research Software*, *4*(1), e25.

Helsel, D. R., and R. M. Hirsch (1993), *Statistical Methods in Water Resources*, 522 pp., Elsevier.

Hendricks, E. L. (1964), Summary of Floods in the United States During 1956, *USGS Water-Supply Paper 1530*, U.S. Geological Survey, Washington D. C.

Herschy, R. (2003), World catalogue of maximum observed floods, *IAHS-AISH Publication 284 320 pp.*, Wallingfor.

Hicks, N. S., J. A. Smith, A. J. Miller, and P. A. Nelson (2005), Catastrophic flooding from an orographic thunderstorm in the central Appalachians, *Water Resources Research*, *41*(W12428), doi:10.1029/2005WR004129.

Hirsch, R. M., and S. A. Archfield (2015), Flood trends: Not higher but more often, *Nature Climate Change*, *5*, 198 – 199.

Hirschboeck, K. K. (1987), Hydroclimatically defined mixed distributions in partial duration flood series, in *Hydrologic Frequency Modelling*, edited by V. P. Singh, pp. 199–212, D. Reidel Publishing Company.

Hirschboeck, K. K. (1988), Flood hydroclimatology, in *Flood Geomorphology*, edited by V. R. Baker, R. C. Kockel, and P. C. Patton, pp. 27–49, John Wiley & Sons.

Hitchens, N. M., and H. E. Brooks (2013), Preliminary investigation of the contribution of supercell thunderstorms to the climatology of heavy and extreme precipitation in the US, *Atmospheric Research*, *123*(206 - 210).

Hosking, J., J. R. Wallis, and E. F. Wood (1985), An Appraisal of the Regional Flood Frequency Procedure in the UK Flood Studies Report, *Hydrological Sciences Journal*, *30*(1), 85 – 109.

Hosking, J. R. M., and J. R. Wallis (1997), *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press.

Jarrett, R. D., and J. E. Costa (1988), Evaluation of the flood hydrology in the Colorado Front Range using precipitation, streamflow and paleoflood data, *USGS Water Resources Investigation Report 87-4177*, U.S. Geological Survey, Washington D. C.

Jarrett, R. D., and E. M. Tomlinson (2000), Regional interdisciplinary paleoflood approach to assess extreme flood potential, *Water Resources Research*, *36*(10), 2957–2984.

Javier, J. R. N., J. A. Smith, J. England, M. L. Baeck, M. Steiner, and A. A. Ntelekos (2007), Climatology of extreme rainfall and flooding from orographic thunderstorm systems in the upper Arkansas River basin, *Water Resources Research*, *43*(W10410), doi:10.1029/2006WR005093.

Jones, J. A., and G. E. Grant (1996), Peakflow responses to clear-cutting and roads in small and large basins, *Water Resources Research*, *32*, 959 – 974.

Jordan, P. R., and L. J. Combs (1996), Summary of Floods in the United States During 1990 and 1991, *Water Supply Paper 2474*, U.S. Geological Survey.

Kampf, S. K., and S. J. Burges (2007), A framework for classifying and comparing distributed hillslope and catchment hydrologic models, *Water Resources Research*, *43*, W05,423, doi:10.1029/2006WR005,370.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Advances in Water Resources*, *25*, 1287–1304.

Koenig, T. A., J. L. Bruce, J. E. O'Connor, B. D. McGee, R. R. Holmes, R. Hollins, B. T. Forbes, M. S. Kohn, M. F. Schellekens, Z. W. Martin, and M. C. Peppler (2016), Identifying and Preserving High-Water Mark Data, *Surface-Water Techniques Chapter 24*, U.S. Geological Survey.

Konrad, C. E. (2001), The most extreme precipitation events over the eastern United States from 1950 to 1996: Considerations of scale, *Journal of Hydrometeorology*, *2*, 309–325.

Koutsoyiannis, D. (1999), A probabilistic view of Hershfield's method for estimating probable maximum precipitation, *Water Resources Research*, *35*(4), 1313 – 1322, w05429, doi:10.1029/2006WR005592.

Kumjian, M. R. (2013), Principles and applications of dual-polarization radar, *Journal of Operational Meteorology*, *1*(19 - 21), 226 – 274.

Leadbetter, M. R., G. Lindgren, and H. Rootzen (1982), *Extremes and Related Properties of Random Sequences and Processes*, Springer Verlag, New York.

Leopold, L. B. (1942), Areal extent of intense rainfalls, New Mexico and Arizona, *Transactions of the American Geophysical Union*, *23*, 558 – 563.

Leopold, L. B. (1946), Two intense local floods in New Mexico, *Transactions of the American Geophysical Union*, *27*(4), 535 – 539.

Levish, D. R., and D. A. Ostenaa (1996), Applied Paleoflood Hydrology in North-Central Oregon, *Guidebook for Fieldtrip 2, April 19-21, 1996, Cordilleran Section 92nd Annual Meeting, GSA*, Bureau of Reclamation, Denver, CO.

Li, C., G. A. Wang, and R. R. Li (2013), Maximum observed floods in China, *Hydrological Sciences Journal*, *58*, 728 – 735.

Liu, M., and J. A. Smith (2016), Extreme Rainfall from Landfalling Tropical Cyclones in the Eastern United States: Hurricane Irene (2011), *Journal of Hydrometeorology*, *17*, 2883 – 2904.

Llasat, M. C., R. Marcos, M. Turco, J. Gilabert, and M. Llasat-Botija (2016), Trends in flash flood events versus convective precipitation in the Mediterranean region: The case of Catalonia, *Journal of Hydrology*, *541*(A), 24 – 37.

Maddox, R. A. (1980), Mesoscale Convective Complexes, *Bulletin of the American Meteorological Society*, *61*(11), 1374 – 1387.

Maddox, R. A., L. R. Hoxit, C. F. Chappell, and F. Caracena (1978), Comparison of meteorological aspects of the Big Thompson and Rapid City flash floods, *Monthly Weather Review*, *106*, 375–389.

Mahoney, K., M. A. Alexander, G. Thompson, J. J. Barsugli, and J. D. Scott (2012), Changes in hail and flood risk in high-resolution simulations over Colorado's mountains, *Nature Climate Change*, *2*, 125 – 131.

Mallakpour, I., and G. Villarini (2015), The changing nature of flooding across the central United States, *Nature Climate Change*, *5*, 250 – 254.

Markowski, P. M., and N. Dotzek (2011), A numerical study of the effects of orography on supercells, *Atmospheric Research*, *100*(4), 457 – 478.

Martins, E. S., and J. R. Stedinger (2000), Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resources Research*, *36*(3), 737–744.

McAnelly, R. L., and W. R. Cotton (1989), The precipitation life cycle of mesoscale convective complexes over the Central United States, *Monthly Weather Review*, *117*, 784–808.

Michaelis, A. C., and G. M. Lackmann (2013), Numerical modeling of a historic storm: Simulating the blizzard of 1888, *Geophysical Research Letters*, *40*, 4092–4097.

Michaud, J. D., K. K. Hirschboeck, and M. Winchel (2001), Regional variations in small-basin floods in the United States, *Water Resources Research*, *37*(1405 - 1416).

Miller, A. J. (1990), Flood hydrology and geomorphic effectiveness in the central Appalachians, *Earth Surface Processes and Landforms*, *15*, 119–134.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity is dead: Whither water management?, *Science*, *319*, 573–574.

Morin, E., D. C. Goodrich, R. A. Maddox, X. G. Gao, H. V. Gupta, and S. Sorooshian (2006), Spatial patterns in thunderstorm rainfall events and their coupling with watershed hydrological response, *Advances in Water Resources*, *29*(6), 843–860, doi:10.1016/j.advwatres.2005.07.014.

Morrison, J. E., and J. A. Smith (2002), Stochastic modeling of flood peaks using the generalized extreme value distribution, *Water Resources Research*, *38*(12), 1305, doi:10.1029/2001WR000502.

Murphy, E. C. (1904), Destructive Floods in the United States in 1903, *Water Supply and Irrigation Paper 96*, U. S. Geological Survey, Washington D. C.

Myers, V. A. (1967), Meteorological estimation of extreme precipitation for spillway design floods, *Tech. Rep. WBTM Hydro-5*, U. S. Weather Bureau, Office of Hydrology, Washington D. C.

Neiman, P. J., L. J. Schick, F. M. Ralph, M. Hughes, and G. A. Wick (2011), Flooding in Western Washington: The Connection to Atmospheric Rivers, *Journal of Hydrometeorology*, *12*(6), 1337 – 1358.

Nguyen, C. C., E. Gaume, and O. Payrastre (2013), Regional flood frequency analyses involving extraordinary flood events at ungauged sites: Further developments and validations, *Journal of Hydrology*, p. DOI 0.1016/j.jhydrol.2013.09.058.

Nielsen, E. R., G. R. Herman, R. C. Tournay, J. M. Peters, and R. S. Schumacher (2015), Double impact: when both tornadoes and flash floods threaten the same place and time, *Weather and Forecasting*, *30*, 1673 – 1693.

Nielsen, E. R. and R. S. Schumacher (2018), Dynamical Insights into Extreme Short-Term Precipitation Associated with Supercells and Mesovortices, *Journal of the Atmospheric Sciences*, in press.

Nielsen-Gammon, J. W., F. Zhang, A. M. Odins, and B. Myoung (2005), Extreme rainfall in Texas: patterns and predictability, *Physical Geography*, *26*(5), 340–364.

O'Connor, J. E., and V. R. Baker (1992), Magnitudes and implications of peak discharges from glacial Lake Missoula, *Geological Society of America Bulletin*, *104*(3), 267 – 279.

O'Connor, J. E., and J. E. Costa (2004a), Spatial distribution of the largest rainfall-runoff floods from basins between 2.6 and 26,000 km^2 in the United States and Puerto Rico, *Water Resources Research*, *40*(W01107), doi:10.1029/2003WR002247.

O'Connor, J. E., and J. E. Costa (2004b), The world's largest floods, past and present - Their causes and magnitude, *Circular 1254*, U.S. Geological Survey.

O'Connor, J. E., G. E. Grant, and J. E. Costa (2002), The geology and geography of floods, *Ancient Floods, Modern Hazards*, *5*, 359 – 385.

Ogden, F. L., H. O. Sharif, S. U. S. Senarath, J. A. Smith, M. L. Baeck, and J. R. Richardson (2000), Hydrologic analysis of the Fort Collins, Colorado, flash flood of 1997, *Journal of Hydrology*, *228*, 82–100.

Parajka, J., et al. (2010), Seasonal characteristics of flood regimes across the Alpine-Carpathian range, *Journal of Hydrology*, *394*, 78 – 89.

Patton, P. C., and V. R. Baker (1976), Morphometry and floods in small drainage basins subject to diverse hydrogeomorphic controls, *Water Resources Research*, *12*(5), 941–952.

Petersen, W. A., L. D. Carey, S. A. Rutledge, J. C. Knievel, N. J. Doesken, R. H. Johnson, T. B. McKee, T. Vonder Haar, and J. F. Weaver (1999), Mesoscale and radar observations of the Fort Collins flash flood of 28 July 1997, *Bulletin of the American Meteorological Society*, *80*(2), 191–216.

Potter, K. W., and J. F. Walker (1981), A model of discontinuous measurement error and its effects on the probability distribution of flood discharge measurements, *Water Resources Research*, *17*(5), 1505–1509.

Robinson, J. S., and M. Sivapalan (1997), An investigation into the physical causes of scaling and heterogeneity of regional flood frequency, *Water Resources Research*, *33*(5), 1045–1059.

Rogash, J. A., and F. Racy (2002), Some Meteorological Characteristics of Significant Tornado Events Occurring in Proximity to Flash Flooding, *Weather and Forecasting*, *17*, 155 – 159.

Romine, R. S., D. W. Burgess, and R. B. Wilhelmson (2008), A Dual-Polarization-Radar-Based Assessment of the 8 May 2003 Oklahoma City Area Tornadic Supercell, *Monthly Weather Review*, *136*, 2849 – 2870.

Rossi, F., M. Fiorentino, and P. Versace (1984), Two-component extreme value distribution for flood frequency analysis, *Water Resources Research*, *20*(7), 847–856.

Rostvedt, J. O. (1970), Summary of Floods in the United States During 1965, *USGS Water-Supply Paper 1850-E*, U.S. Geological Survey, Washington D. C.

Ryberg, K. R., B. B. Goree, T. Williams-Sether, and R. R. Mason Jr. (2017), The U.S. Geological Survey Peak-Flow File Data Verification Project, *Scientific Investigations Report 2017-5119*, U. S. Geological Survey.

Salinas, J. L., A. Castellarin, A. Viglione, S. Kohnova, and T. R. Kjeldsen (2014), Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent?, *Hydrology and Earth System Sciences*, *18*, 4381 – 4389.

Scheffknecht, P., S. Serafin, and V. Grubisic (2017), A long-lived supercell over mountainous terrain, *Quarterly Journal of the Royal Meteorological Society*, *143*, 2973 – 2986.

Schumacher, R. S., and R. H. Johnson (2005), Organization and environmental properties of extreme-rain-producing mesoscale convective systems, *Monthly Weather Review*, *133*(4), 961–976, doi:10.1175/MWR2899.1.

Smith, A., C. Sampson, and P. Bates (2015), Regional flood frequency analysis at the global scale, *Water Resources Research*, *51*, 539 – 553.

Smith, B. K., and J. A. Smith (2015), The flashiest watersheds in the contiguous United States, *Journal of Hydrometeorology*, *16*(12), 2365 – 2381.

Smith, J. A., and M. L. Baeck (2015), "Prophetic Vision, Vivid Imagination": The 1927 Mississippi River Flood, *Water Resources Research*, *51*, 9964 – 9994.

Smith, J. A., M. L. Baeck, M. Steiner, and A. J. Miller (1996), Catastrophic rainfall from an upslope thunderstorm in the Central Appalachians: the Rapidan Storm of June 27, 1995, *Water Resources Research*, *32*(10), 3099–3113.

Smith, J. A., M. L. Baeck, J. E. Morrison, and P. Sturdevant-Rees (2000), Catastrophic rainfall and flooding in Texas, *Journal of Hydrometeorology*, *1*, 5– 25.

Smith, J. A., M. L. Baeck, Y. Zhang, and C. A. Doswell III (2001), Extreme rainfall and flooding from supercell thunderstorms, *Journal of Hydrometeorology*, *2*, 469–489.

Smith, J. A., G. Villarini, and M. L. Baeck (2011a), Mixture distributions and the climatology of extreme rainfall and flooding in the eastern US, *Journal of Hydrometeorology*, *12*(2), 294–309.

Smith, J. A., M. L. Baeck, A. A. Ntelekos, G. Villarini, and M. Steiner (2011b), Extreme rainfall and flooding from orographic thunderstorms in the Central Appalachians, *Water Resources Research*, *47*(W04514), doi:10.1029/2010WR010190.

Soderholm, B., B. Ronalds, and D. N. Kirshbaum (2014), The evolution of convective storms initiated by an isolated mountain ridge, *Monthly Weather Review*, *142*, 1430 – 1451.

Sprenger, M., and H. Wernli (2015), The LAGRANTO Lagrangian analysis tool: Versio 2.0, *Geoscientific Model Development*, *8*, 2569–2586.

Tuttle, J. D., and C. A. Davis (2006), Corridors of warm season precipitation in the central United States, *Monthly Weather Review*, *134*(9), 2297–2317, doi:10.1175/MWR3188.1.

Venables, W. N., and B. D. Ripley (2002), *Modern Applied Statistics with S*, Springer, New York.

Villarini, G. (2016), On the seasonality of flooding across the continental United States, *Advances in Water Resources*, *87*, 80–91.

Villarini, G., and J. A. Smith (2010), Flood peak distributions for the eastern United States, *Water Resources Research*, *46*(W06504), doi:10.1029/2009WR008395.

Villarini, G., F. Serinaldi, J. A. Smith, and W. F. Krajewski (2009), On the stationarity of annual flood peaks in the continental United States during the 20th century, *Water Resources Research*, *45*(W08417), doi:10.1029/2008WR007645.

Villarini, G., J. A. Smith, M. L. Baeck, T. Marchock, and G. A. Vecchi (2011), Characterization of rainfall distribution and flooding associated with U.S. landfalling tropical cyclones: Analyses of Hurricanes Frances, Ivan, and Jeanne (2004), *Journal of Geophysical Research*, *116*, D23,116.

Vogel, R. M., A. Zafirakou-Koulouris, and N. C. Matalas (2001), Frequency of record-breaking floods in the United States, *Water Resources Research*, *37*(6), 1723 – 1731.

Wells, J. V. (1959), Summary of Floods in the United States During 1953, *USGS Water-Supply Paper 1320-E*, U.S. Geological Survey, Washington D. C.

Whistler, J. T. (1903), The Heppner Disaster, *Engineering News Record*, *L*(3), 53 – 54.

Wood, E. F., M. Sivapalan, and K. Beven (1990), Similarity and Scale in Catchment Storm Response, *Reviews of Geophysics*, *28*(1), 1–18, doi:

10.1029/RG028i001p00001.

Woolley, R. R. (1946), Cloudburst Floods in Utah 1850 - 1938, *Water Supply Paper 994*, U.S. Geological Survey, US Government Printing Office, Washinton DC.

Yang, L., M. Liu, J. A. Smith, and F. Tian (2017), Typhoon Nina and the August 1975 flood over Central China, *Journal of Hydrometeorology*, *18*, 451 – 472.

Zhang, Y., J. A. Smith, and M. L. Baeck (2001), The hydrology and hydrometeorology of extreme floods in the Great Plains of Eastern Nebraska, *Advances in Water Resources*, *24*, 1037–1049.

Zhu, L., and S. M. Quiring (2013), Variations in tropical cyclone precipitation in Texas (1950 to 2009), *Journal of Geophysical Research - Atmospheres*, *118*, 3085 – 3096.

Zipser, E. J., D. J. Cecil, C. Liu, S. W. Nesbitt, and D. P. Yorty (2006), Where are the most intense thunderstorms on earth?, *Bulletin of the American Meteorological Society*, *87*, 1057–1071.

Zorzetto, E., G. Botter, and M. Marani (2016), On the emergence of rainfall extremes from ordinary events, *Geophysical Research Letters*, *43*(15), 8076 – 8082.

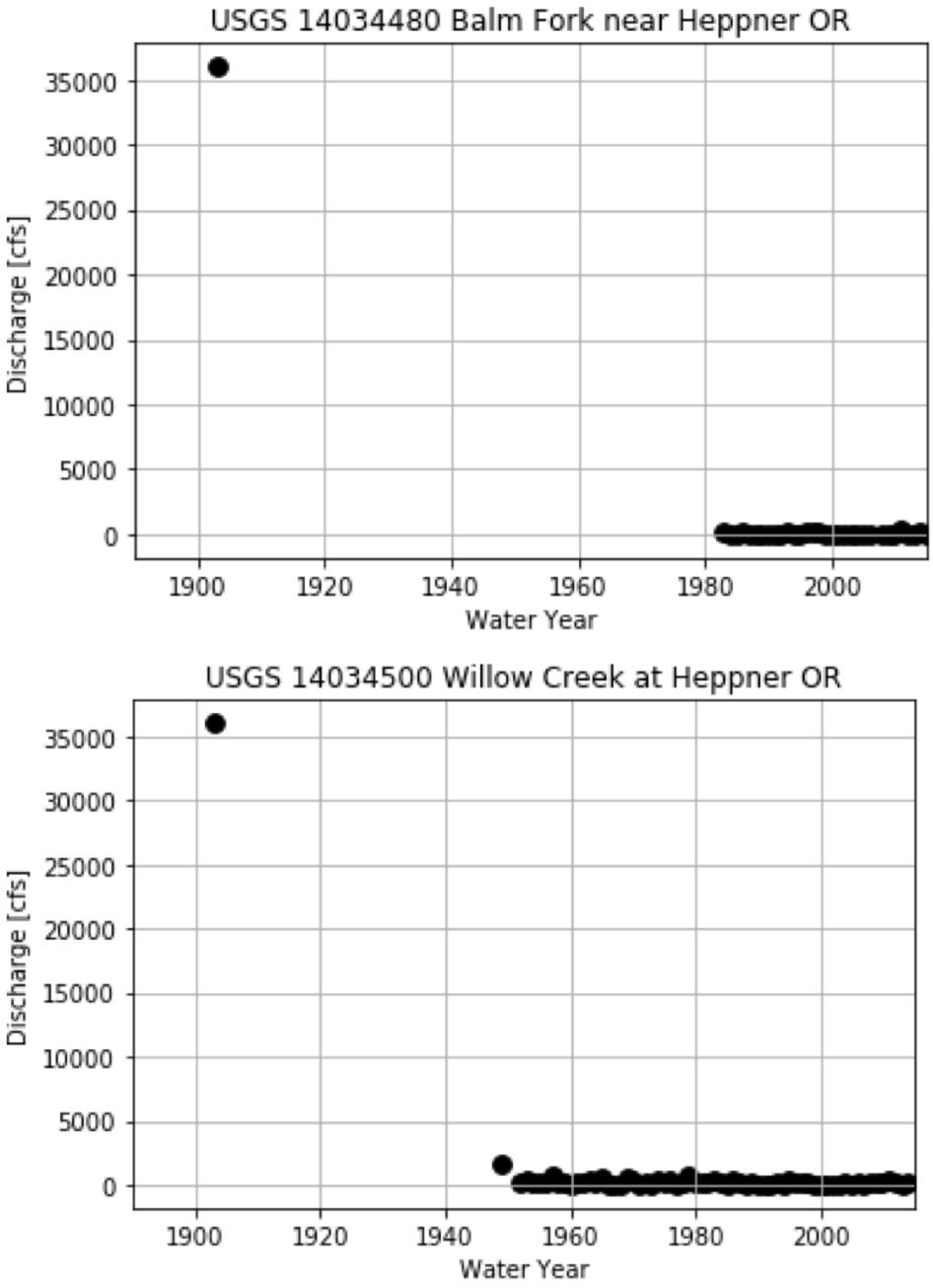


Figure 1. Annual maximum flood peaks for Balm Fork (top) and Willow Creek (bottom). Balm Fork has a drainage area of 68 km^2 and Willow Creek has a drainage area of 251 km^2 (see also Table 1).

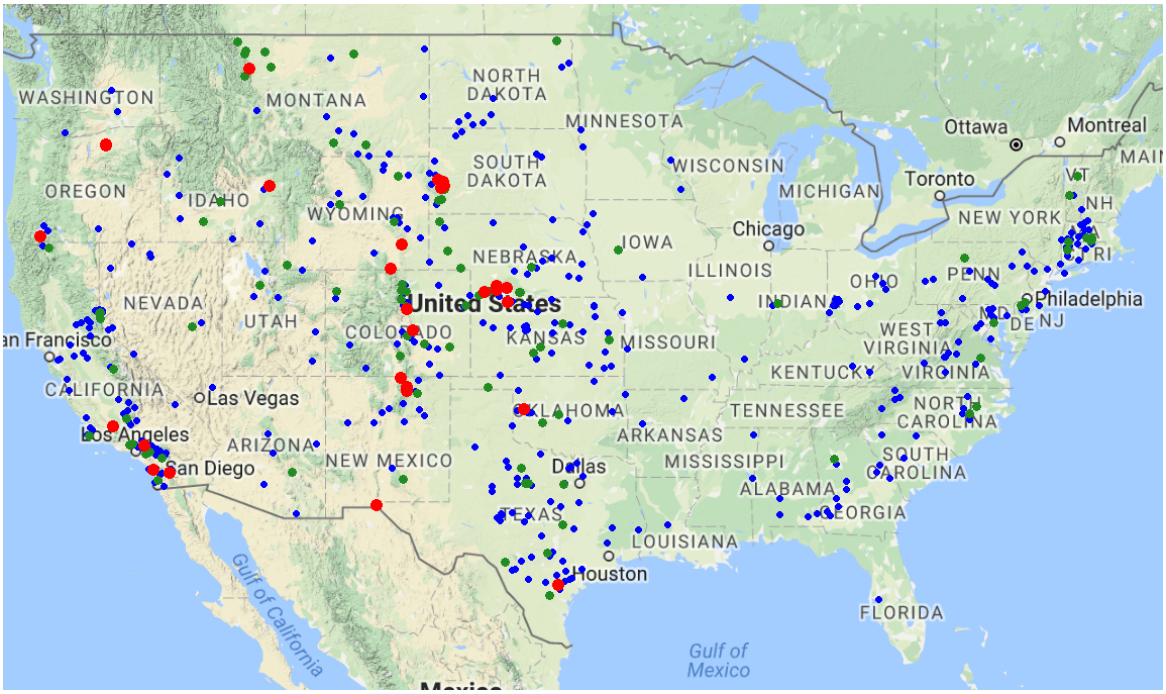
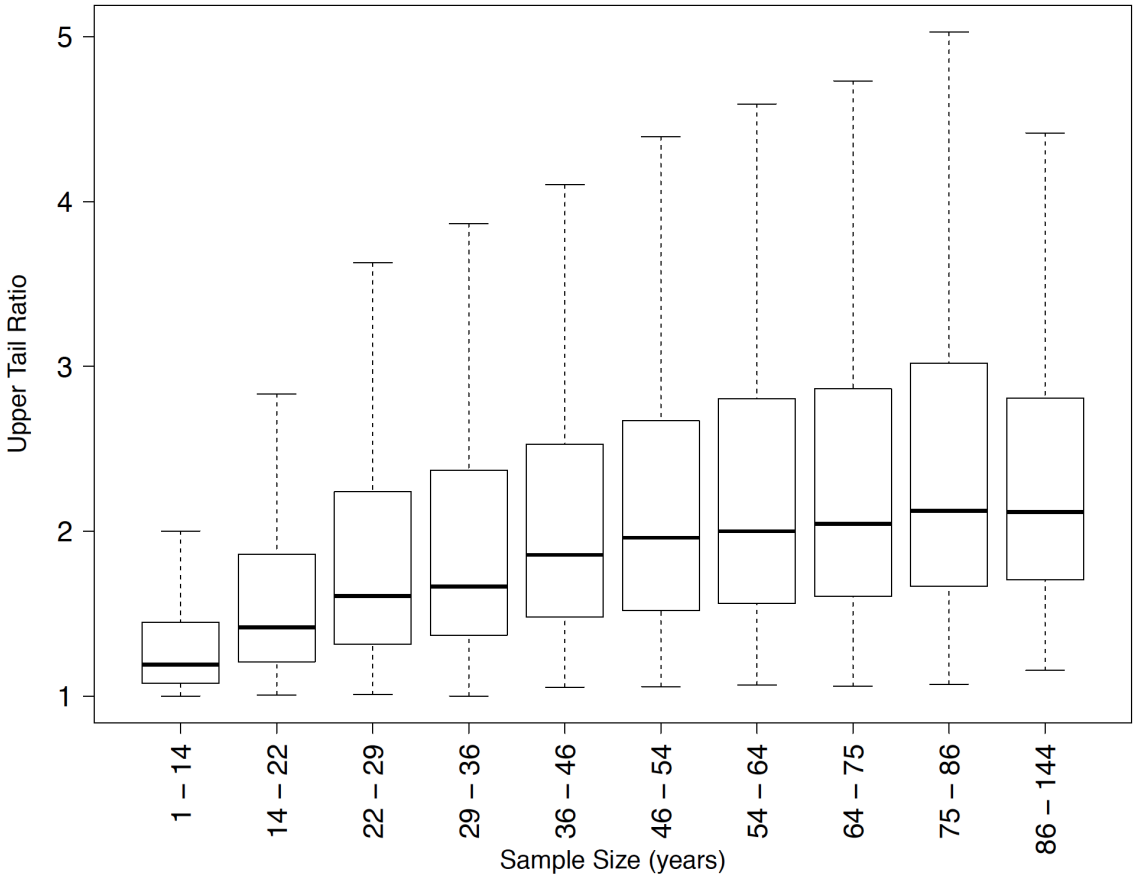


Figure 2. Locations of USGS gaging stations with upper tail ratios exceeding 5 and less than 10 (blue), between 10 and 20 (green) and greater than 20 (red).



1143 **Figure 3.** Boxplots of upper tail ratio organized by deciles of sample size in years. For each
1144 decile, the minimum and maximum values of sample size are shown.

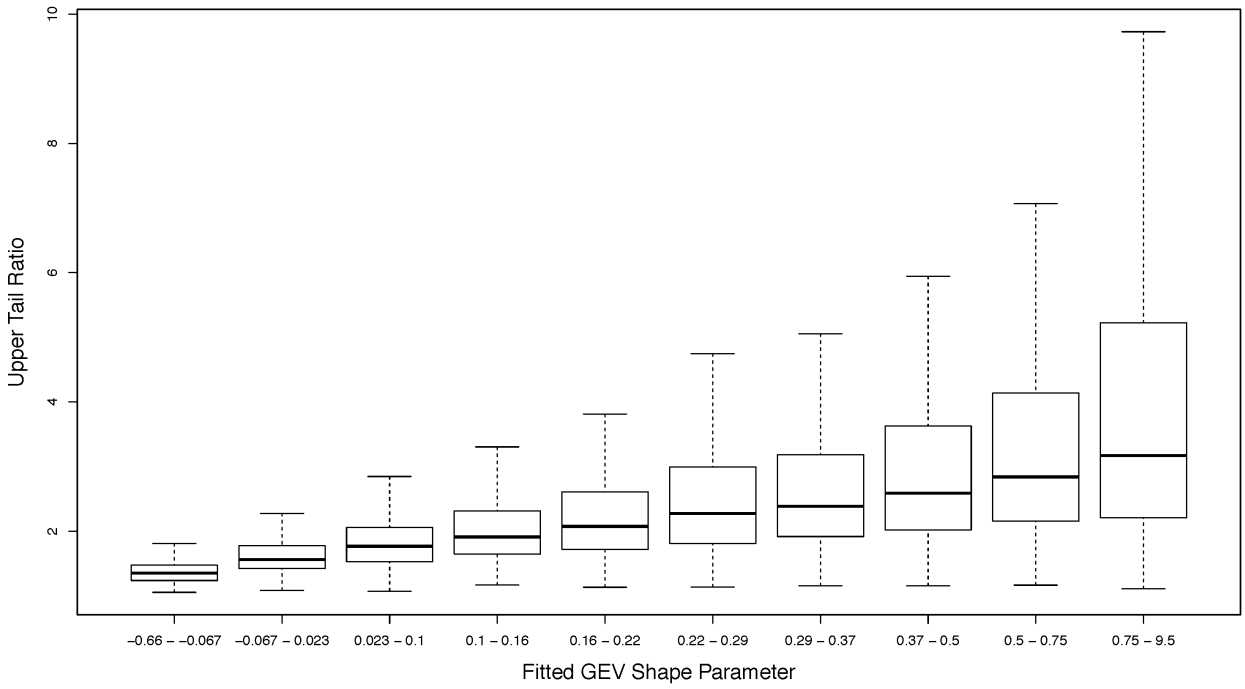


Figure 4. Boxplots of upper tail ratio organized by deciles of estimated GEV shape parameter. For each decile, the minimum and maximum values of sample size are shown.

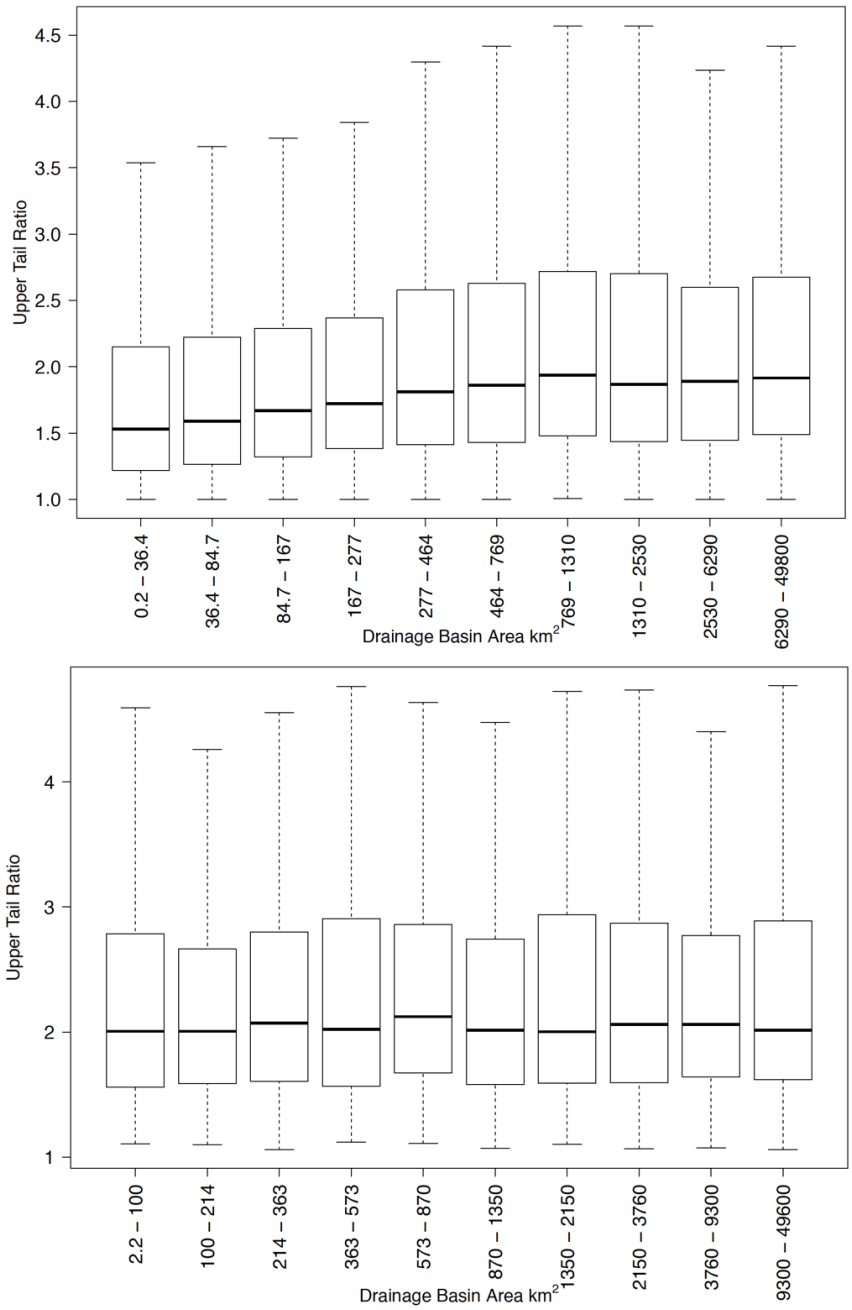


Figure 5. Boxplots of upper tail ratio organized by deciles of basin area in km^2 (top). For each decile, the minimum and maximum values of basin area are shown. In the bottom panel of the figure, the sample of stations is conditioned on sample size exceeding 46 years, which is the median sample size (see Figure 3). Consequently, deciles of basin area in the bottom plot are for the smaller set of stations with sample size greater than 46 years.

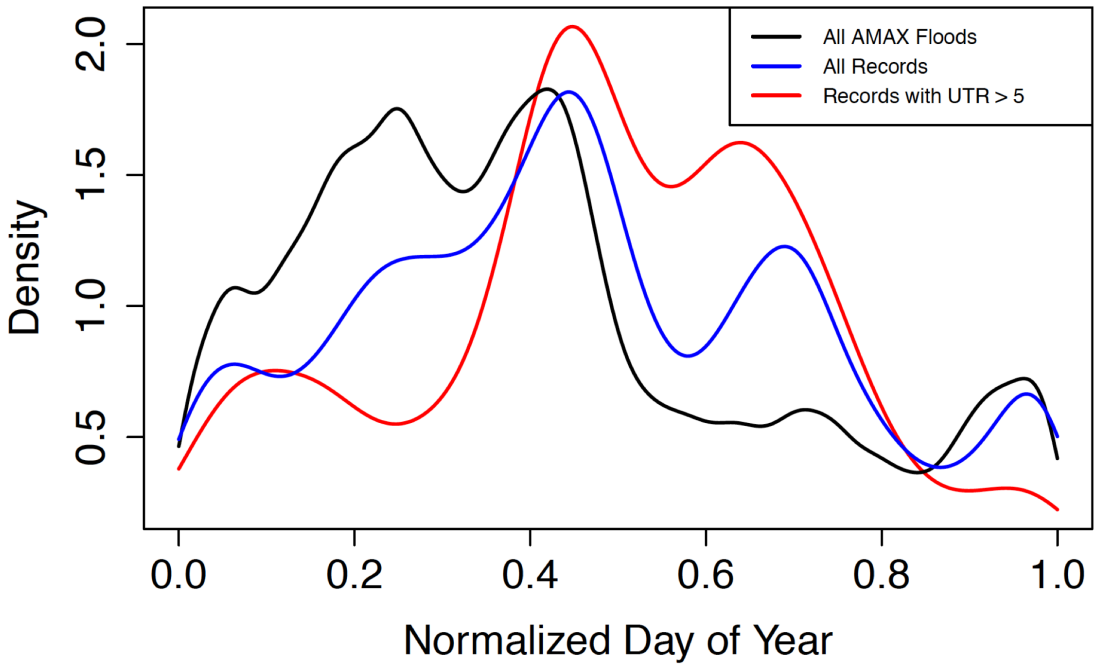


Figure 6. Sample probability density functions (based on R Kernel Density Estimator) of occurrence times of record floods (blue line), records floods with upper tail ratio greater than 5 (red line) and of all flood peaks from the 8911 stream gaging stations (black line; denoted AMAX floods). Time 0 corresponds to January 1 and time 1 to December 31.

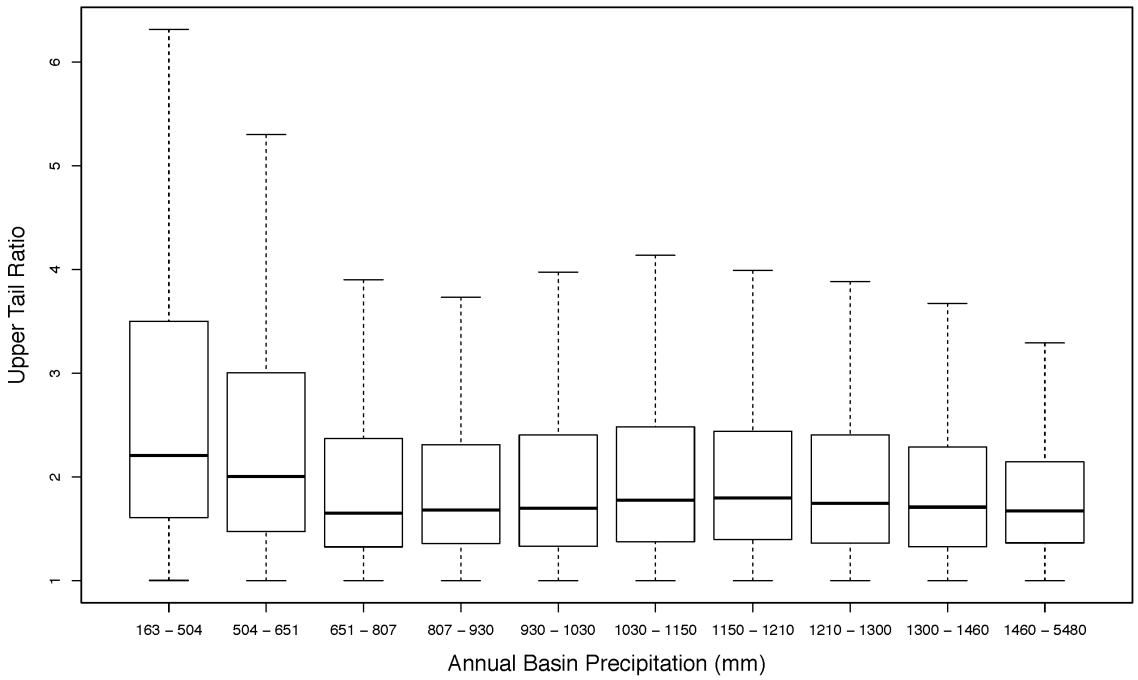
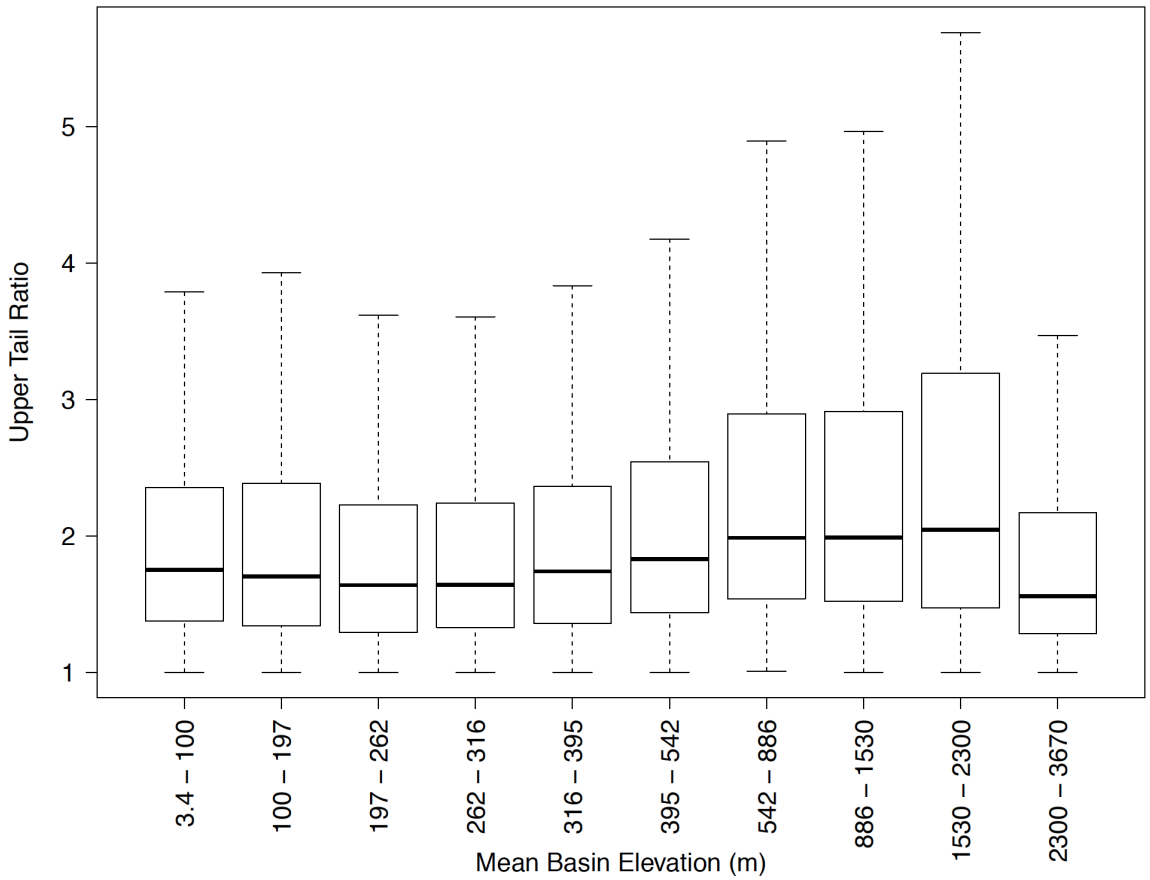
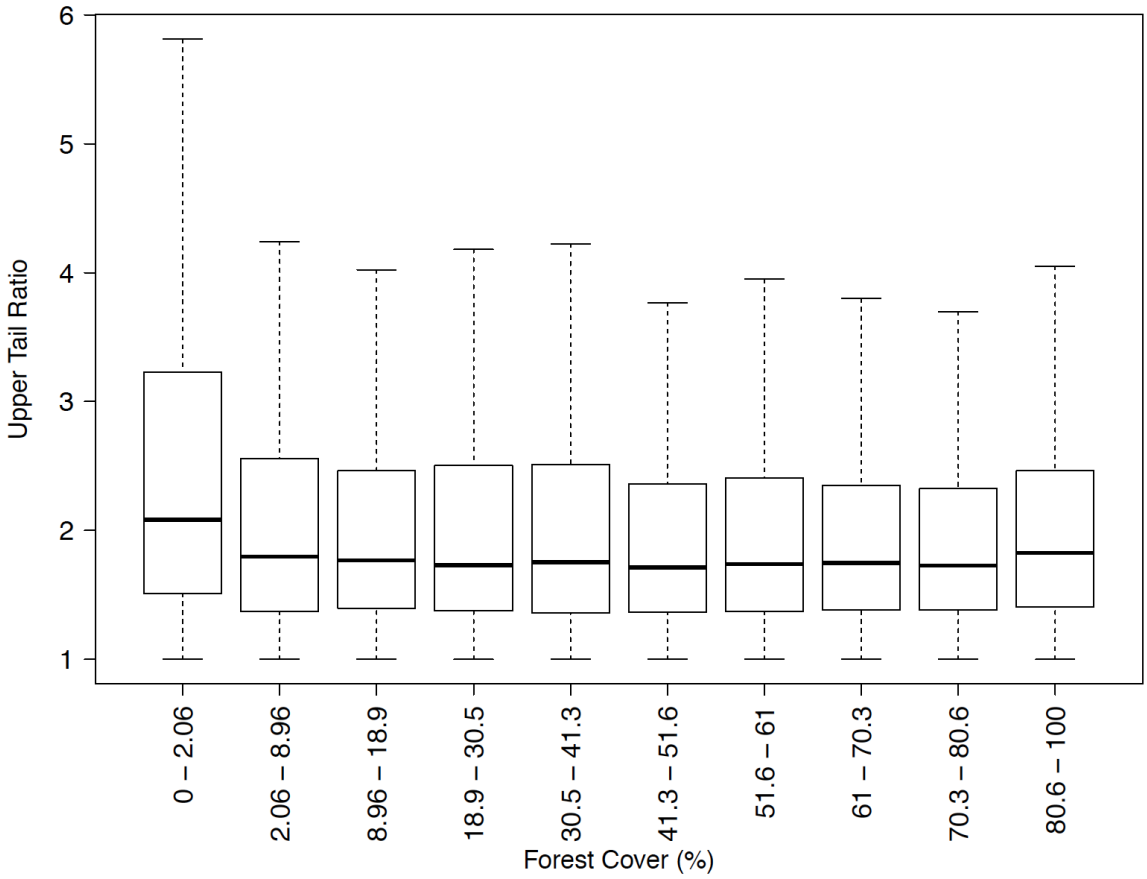


Figure 7. Boxplots of upper tail ratio organized by deciles of mean annual precipitation for the drainage basin (mm). For each decile, the minimum and maximum values of precipitation are shown.



1159 **Figure 8.** Boxplots of upper tail ratio organized by deciles of mean basin elevation (meters
1160 above msl). For each decile, the minimum and maximum values of elevation are shown.



1161 **Figure 9.** Boxplots of upper tail ratio organized by deciles of per cent forest cover. For each
1162 decile, the minimum and maximum values of per cent forest cover are shown.

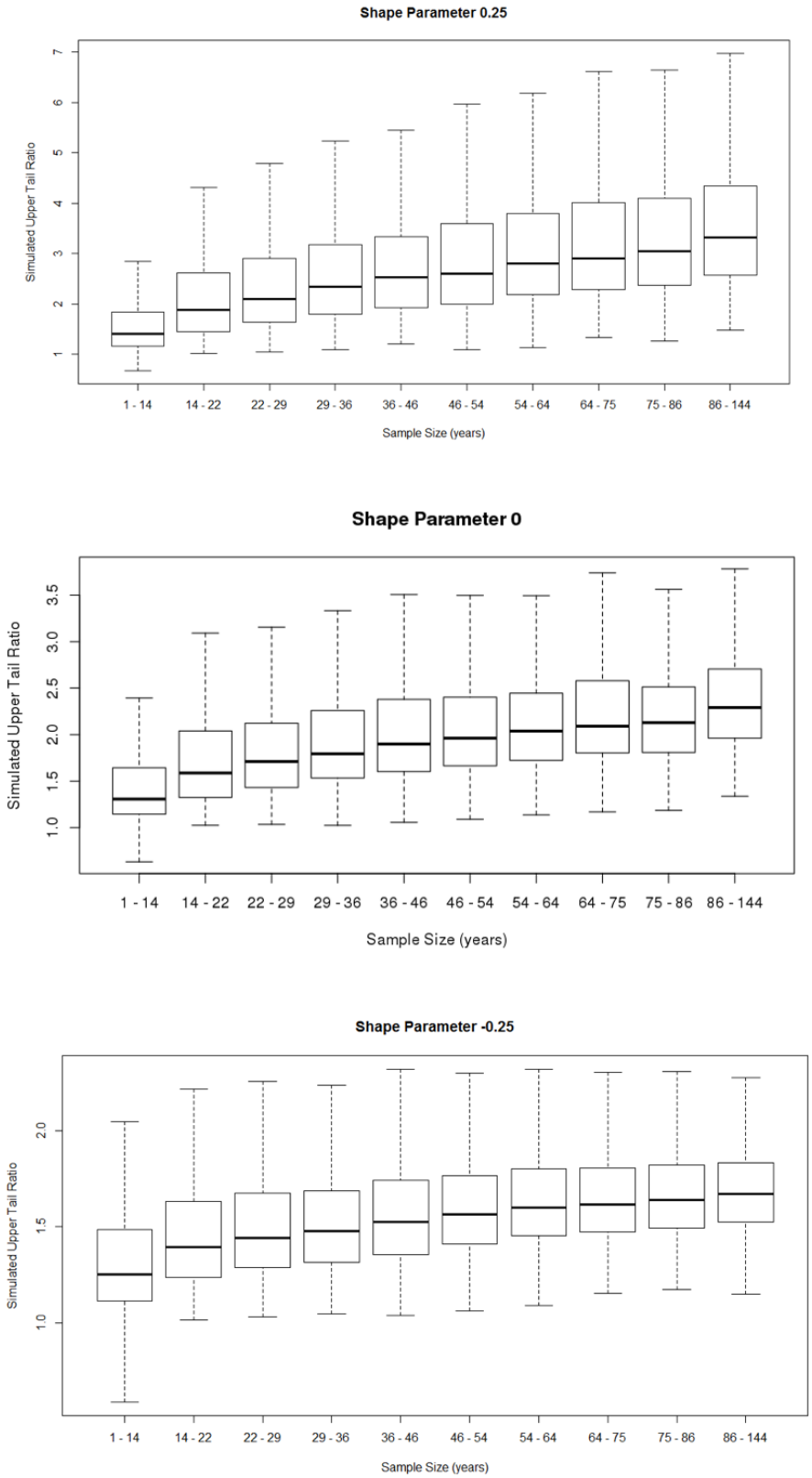
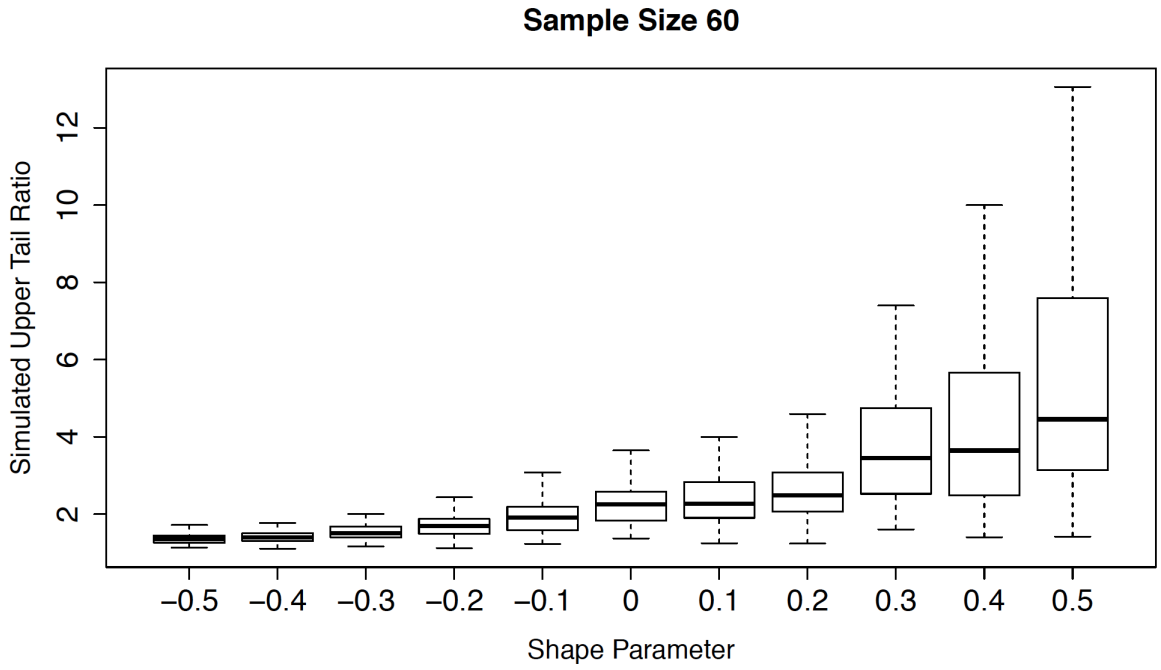


Figure 10. Boxplots of simulated upper tail ratio from GEV distributions with shape parameter 0.25 (top; unbounded, thick-tailed), 0 (middle; Gumbel) and -0.25 (bottom; bounded). The location parameter is 0 and shape parameter 1 for all simulations. Boxplots are organized by deciles of sample size.



1167 **Figure 11.** Boxplots of simulated upper tail ratio from GEV distributions with shape parame-
1168 ters ranging from -0.5 to 0.5 with sample size of 60 years.

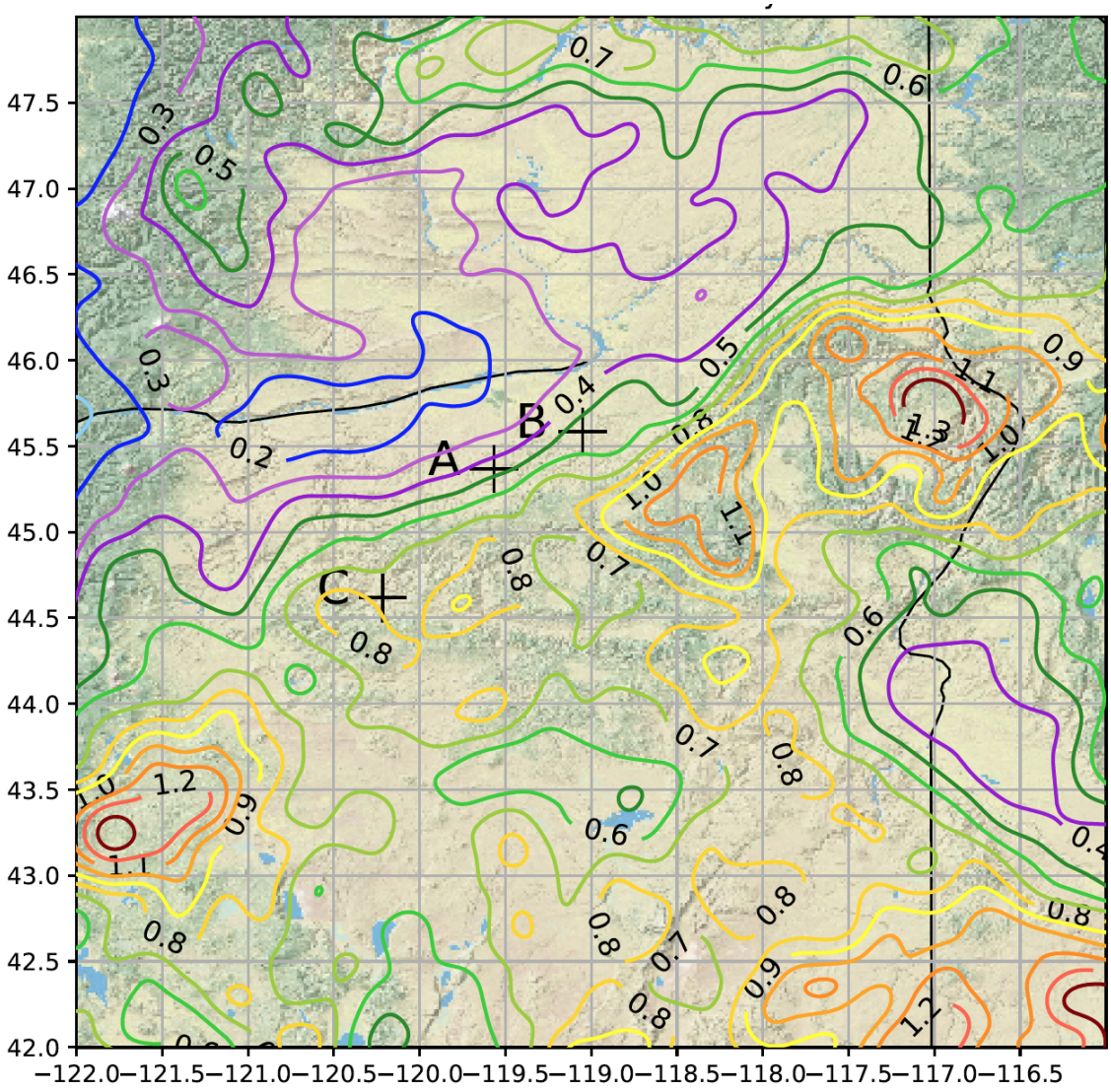
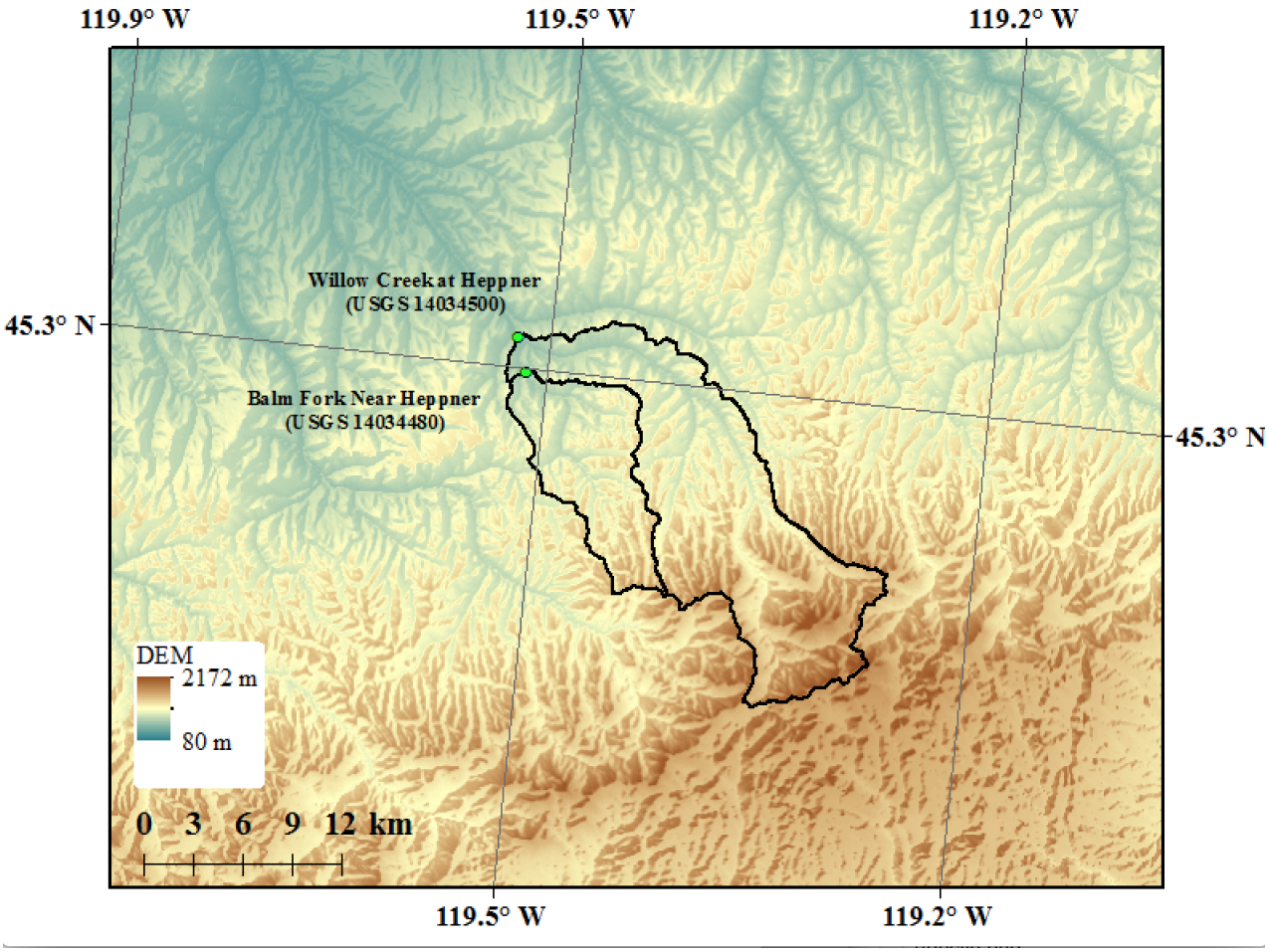
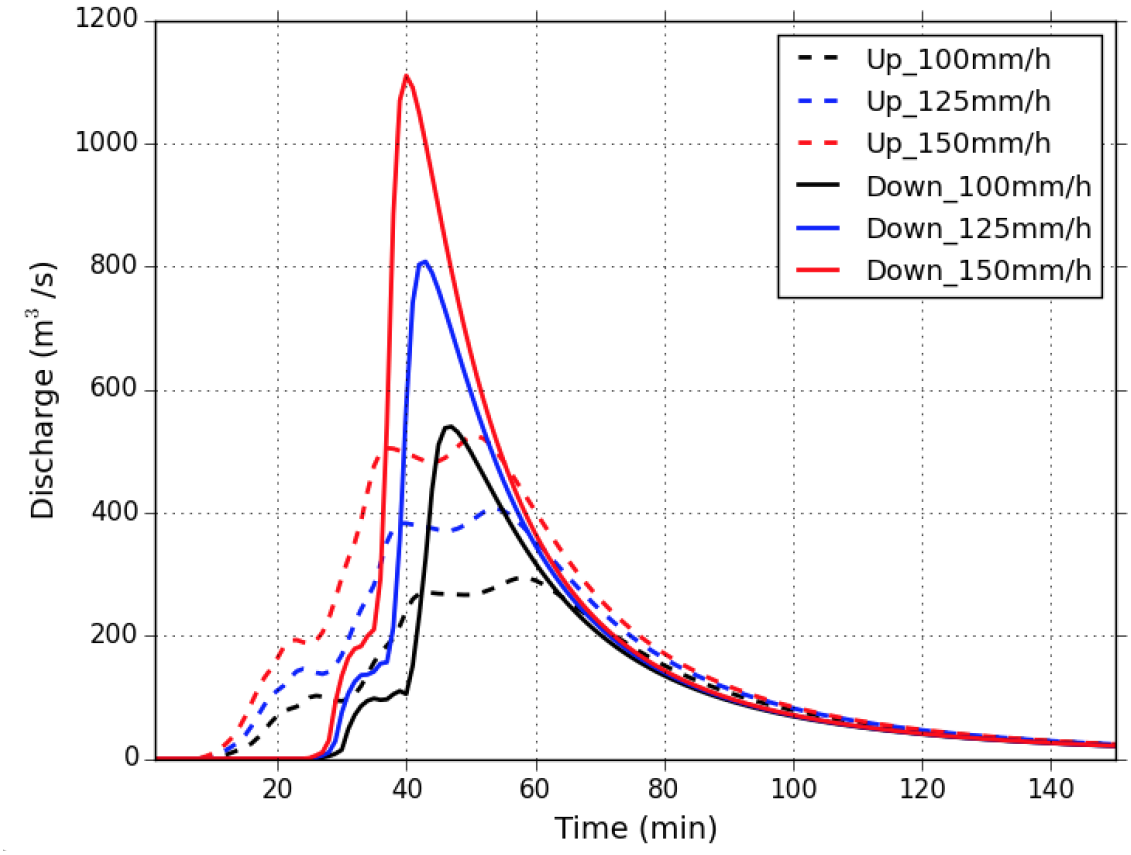


Figure 12. Mean annual cloud-to-ground lightning flash density (CG strikes km^{-2}) for the Blue Mountains region of eastern Oregon. Locations of the Heppner, Lane Canyon and Meyers Canyon floods are denoted by "A", "B" and "C", respectively.



1172 **Figure 13.** Basin boundaries of Balm Fork and Willow Creek, with topography.



1173 **Figure 14.** KINEROS-2 model reconstruction of 14 June 1903 flood in Balm Fork. Results
1174 for rainfall rates of 100, 125 and 150 $mm\ h^{-1}$ for downstream storm motion (solid lines) and
1175 upstream storm motion (dashed lines).

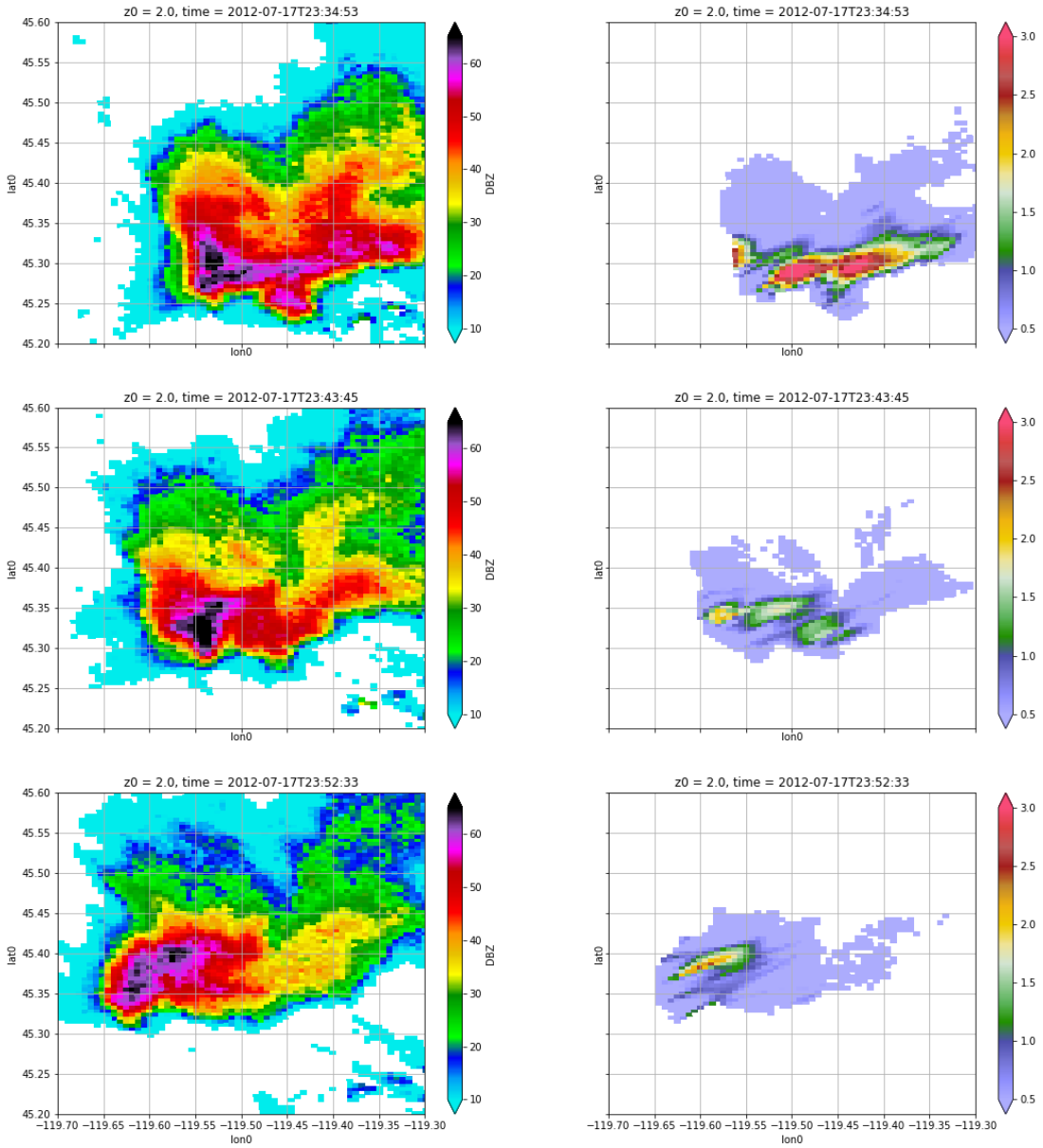


Figure 15. Reflectivity fields (dBZ; left) and specific differential phase shift fields (degrees km^{-1}; right) at 2 km above ground level for the 17 July 2012 storm at 2334 UTC, 2343 UTC and 2352 UTC (prepared using the Python ARM Radar Toolkit; see [*Helmus and Collis*, 2016]).

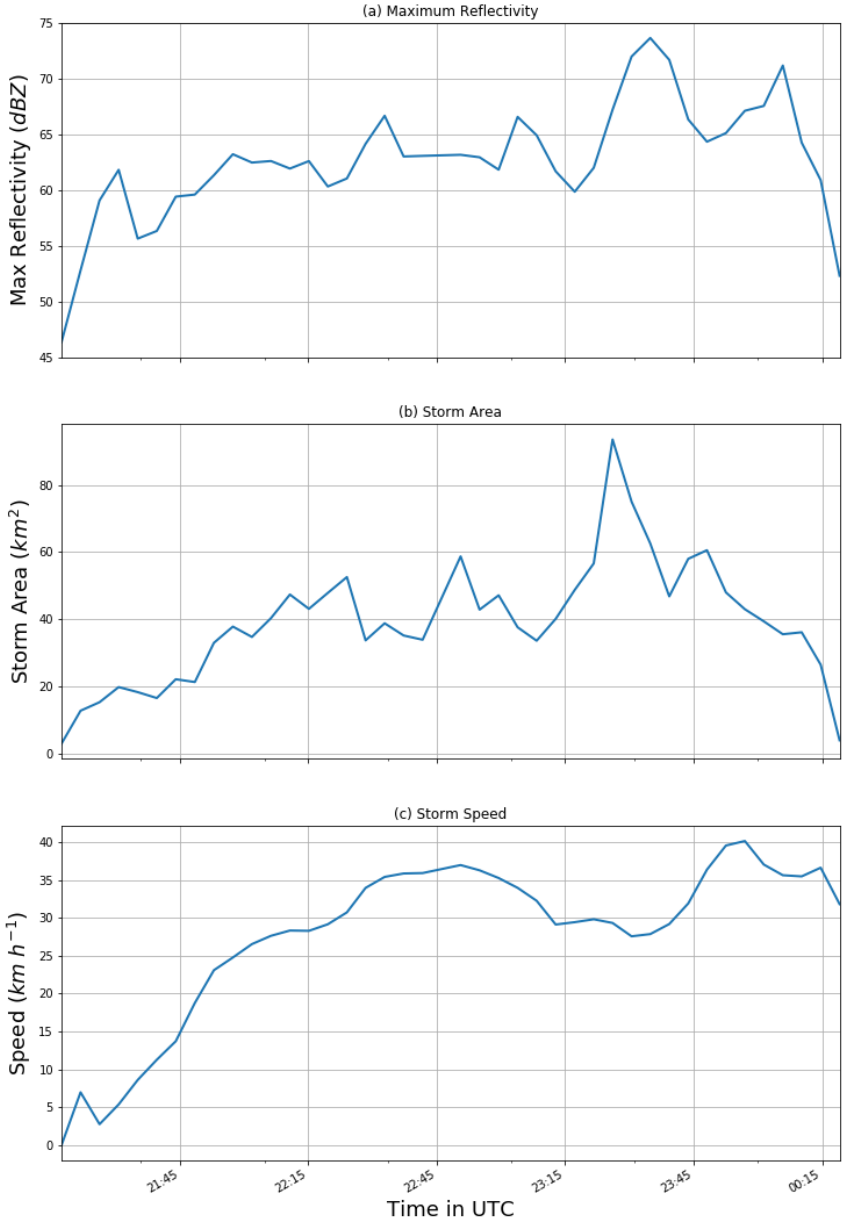
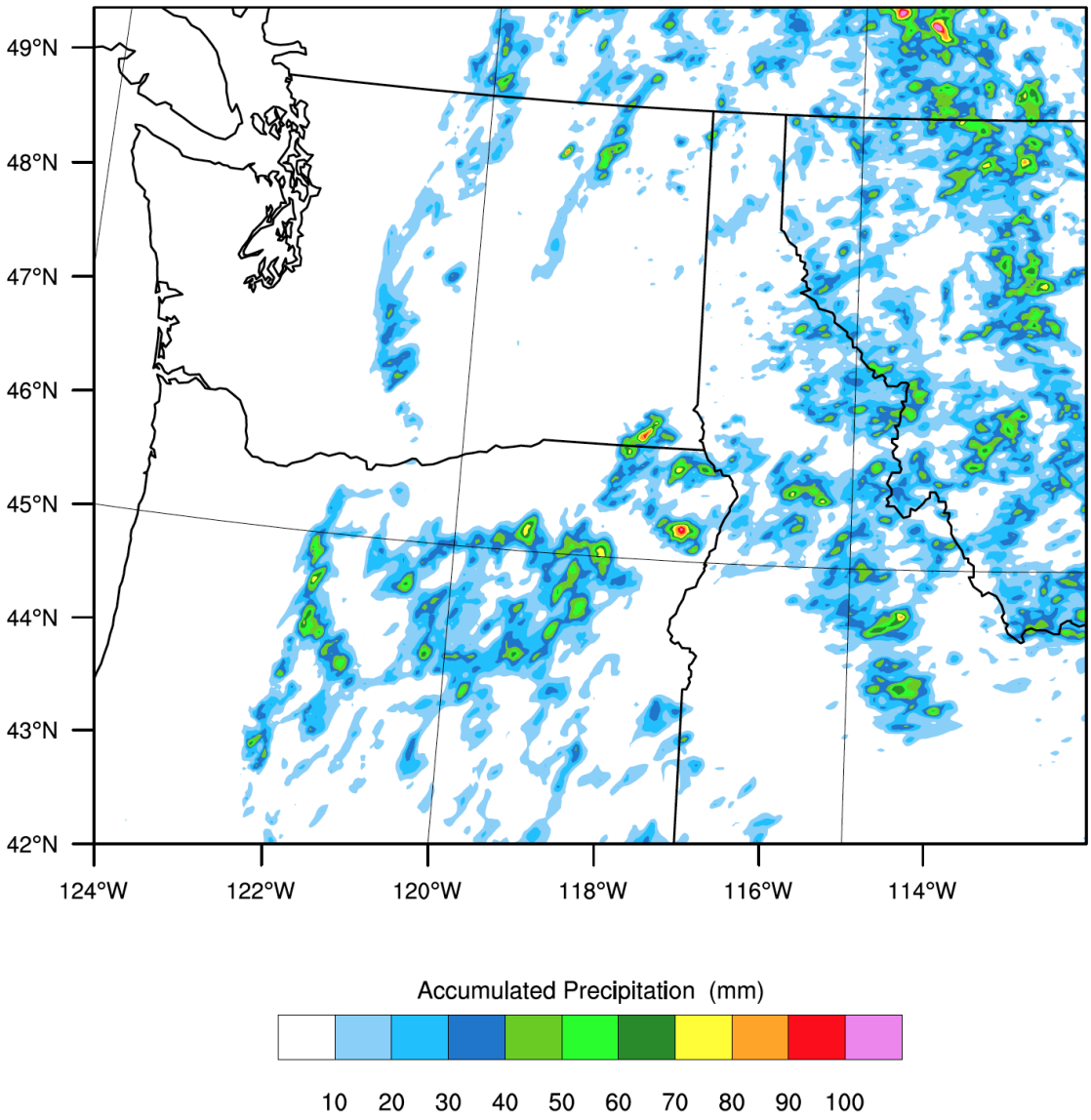


Figure 16. Time series of maximum reflectivity (top; dBZ), storm size (middle; km^2) and storm speed (bottom; $km\ h^{-1}$) for the 17 July 2012 storm (see Figure 15).



1181 **Figure 17.** WRF storm total rain (mm) for the 14 June 1903 Heppner storm.

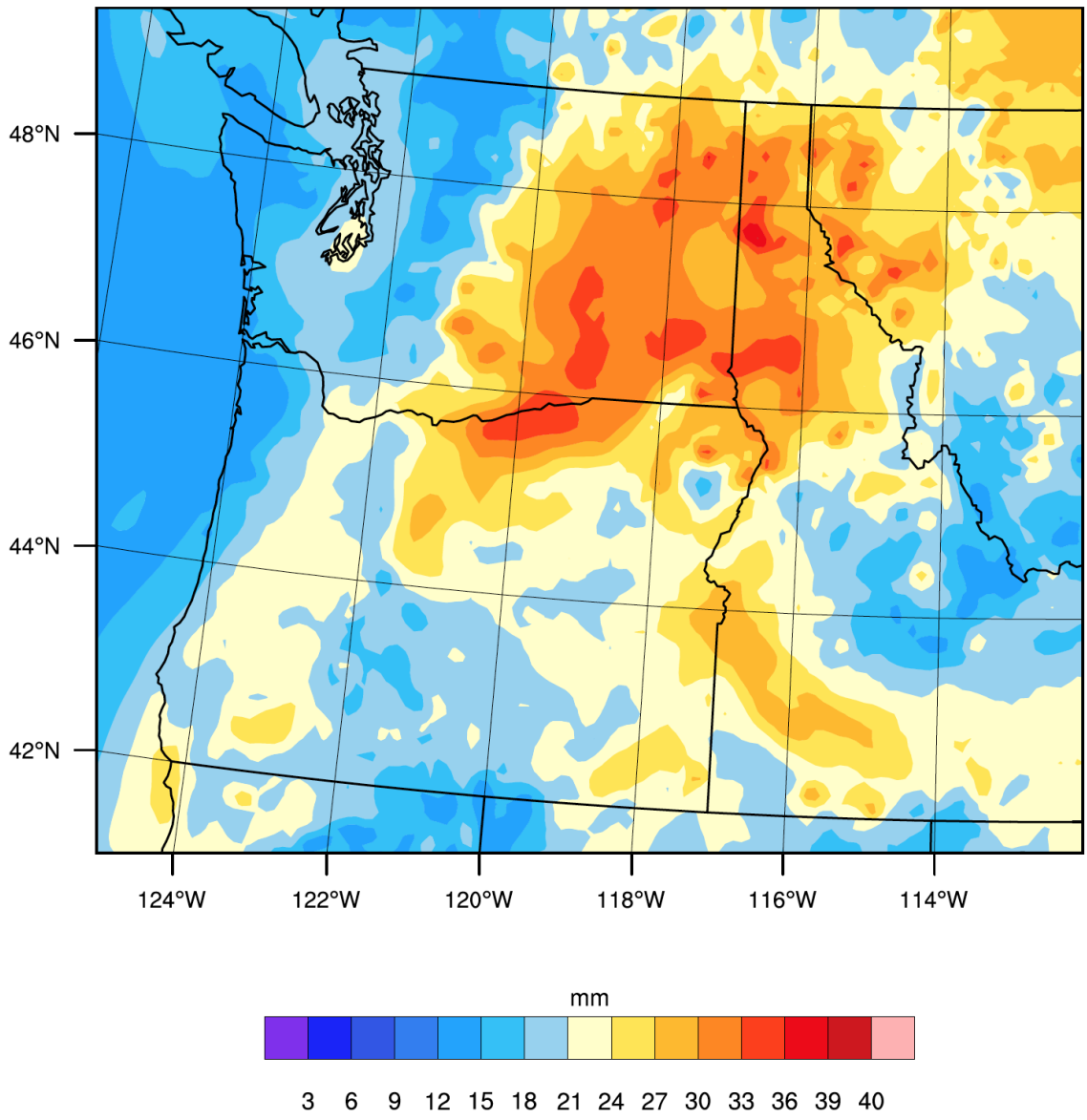
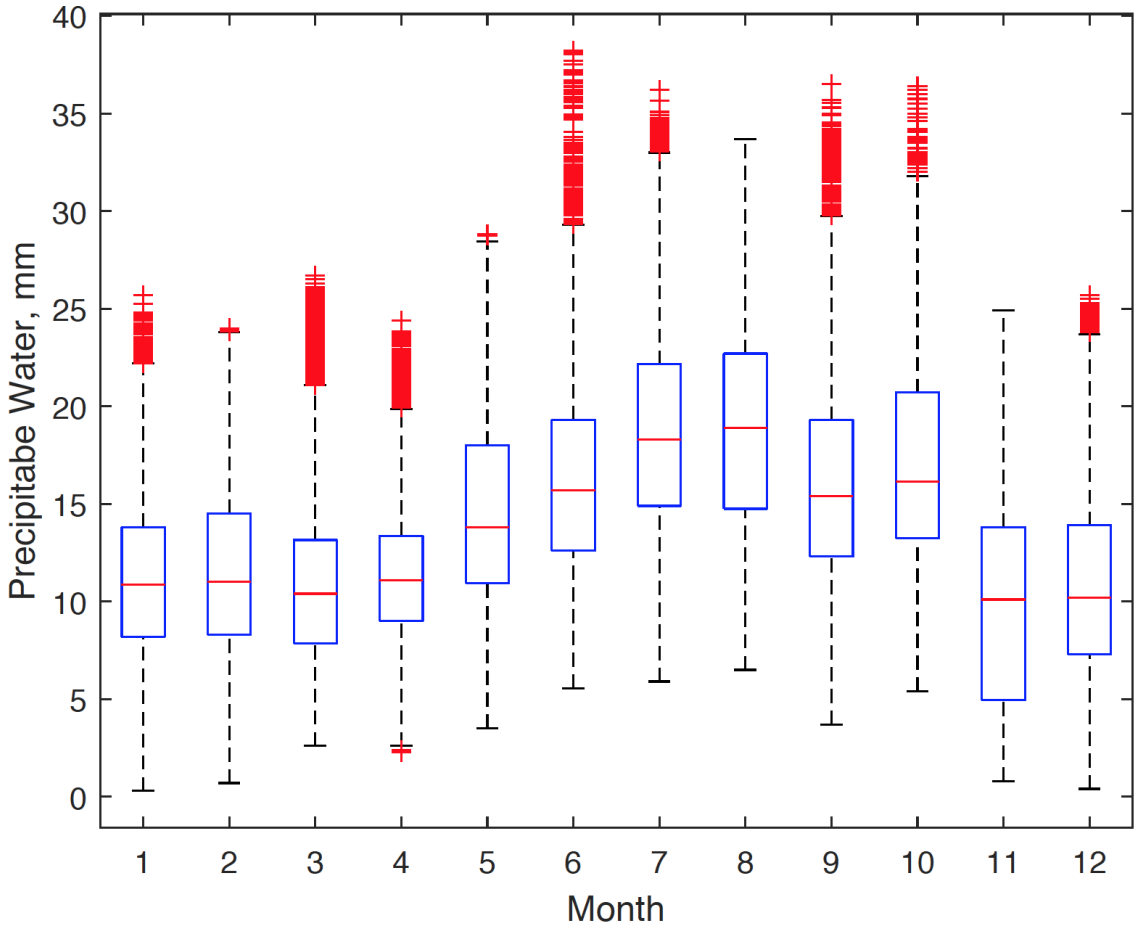


Figure 18. WRF precipitable water (mm) at 2000 UTC on 14 June 1903.



1183 **Figure 19.** Monthly box plots of precipitable water (mm) for Heppner, Oregon based on GPS
1184 precipitable water measurements. Whiskers represent 0.1 and 0.9 quantiles of precipitable water.
1185 Outliers are denoted by red plus signs.

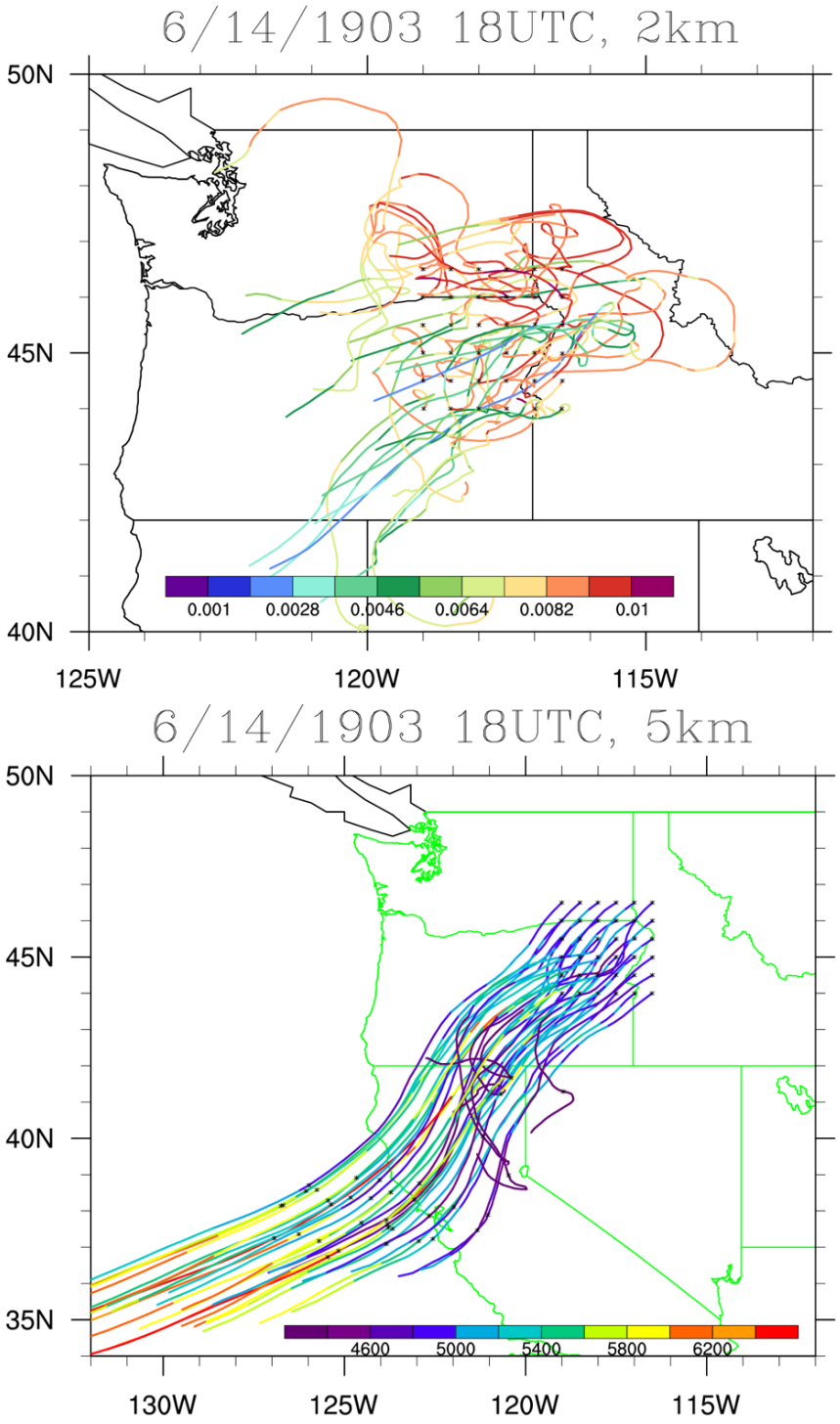


Figure 20. Lagranto back-trajectories for 5km (elevation) and 2 km (specific humidity) ending at 1800 UTC on 14 June 1903 at 2 km elevation (top) and 5 km elevation (bottom). Color coding for the 2km analyses (top) is specific humidity of the parcel ($kg\ kg^{-1}$); color coding for the 5 km analyses (bottom) is elevation (m).

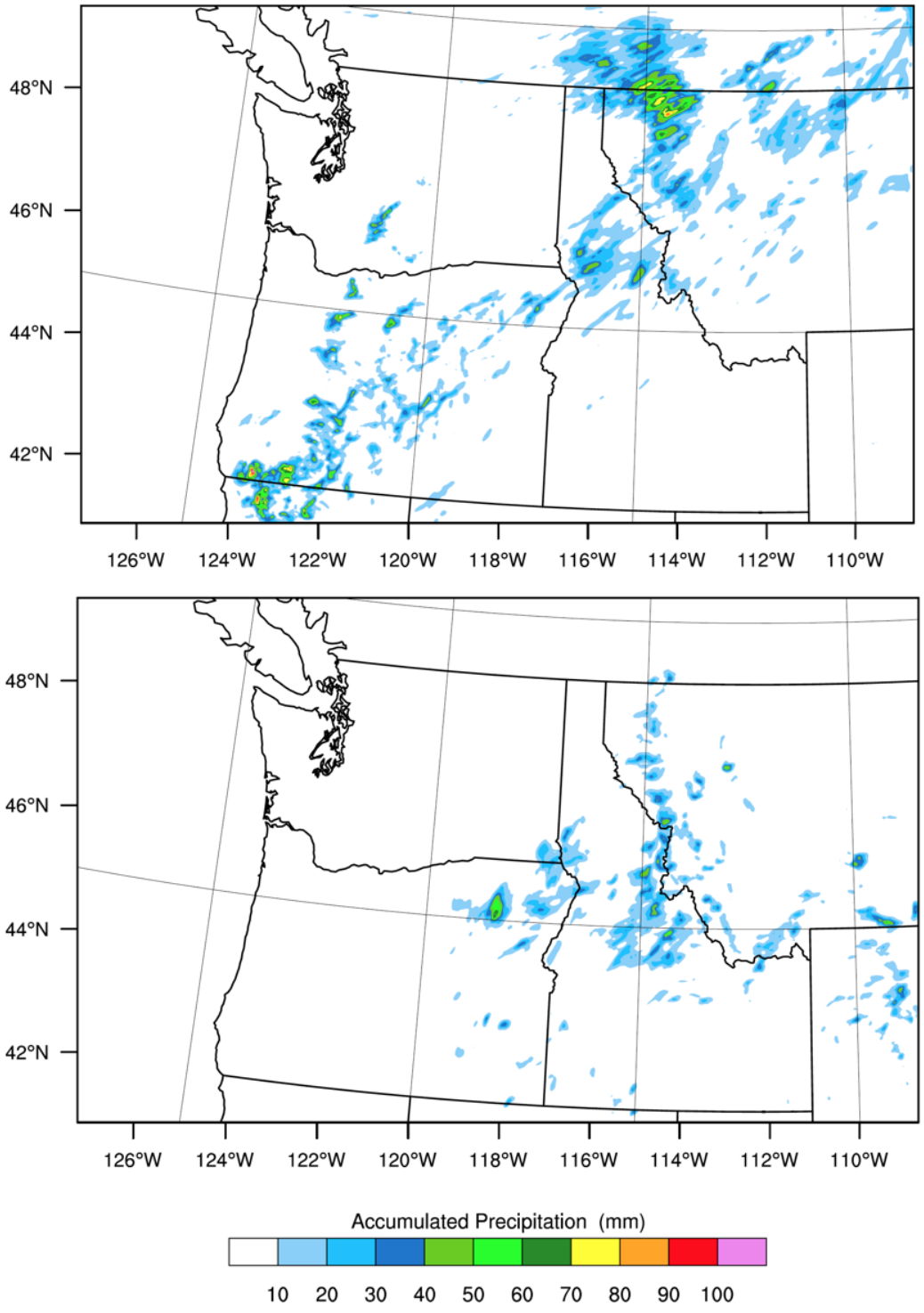


Figure 21. WRF storm total rain (mm) for the 13 July 1956 Meyers Canyon storm (top) and 26 July 1965 Lane Canyon storm (bottom).

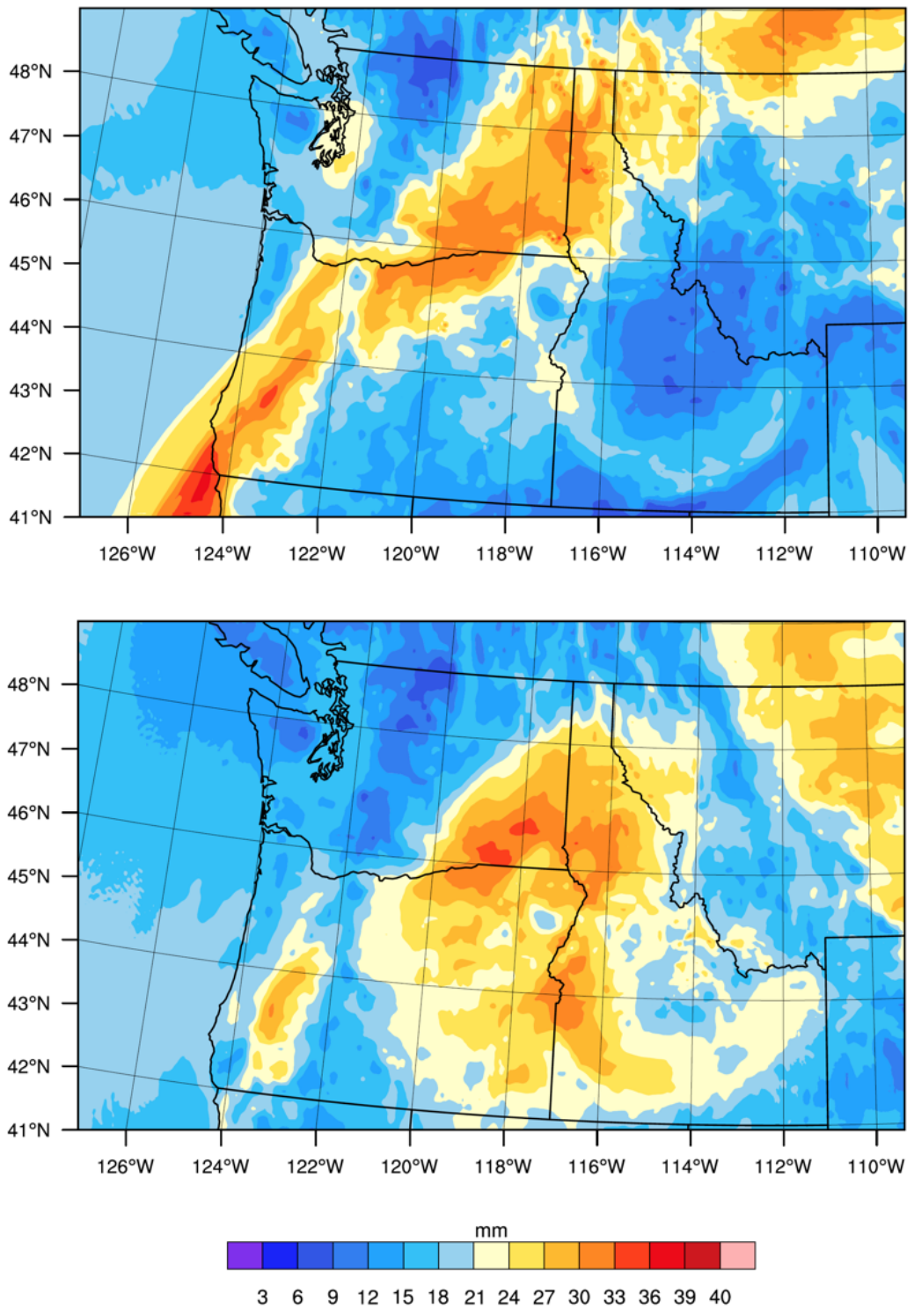


Figure 22. WRF pre-storm precipitable water (mm) for Meyers Canyon storm (13 July 1956 at 1800 UTC; top) and Lane Canyon storm (26 July 1965 at 1800 UTC; bottom).

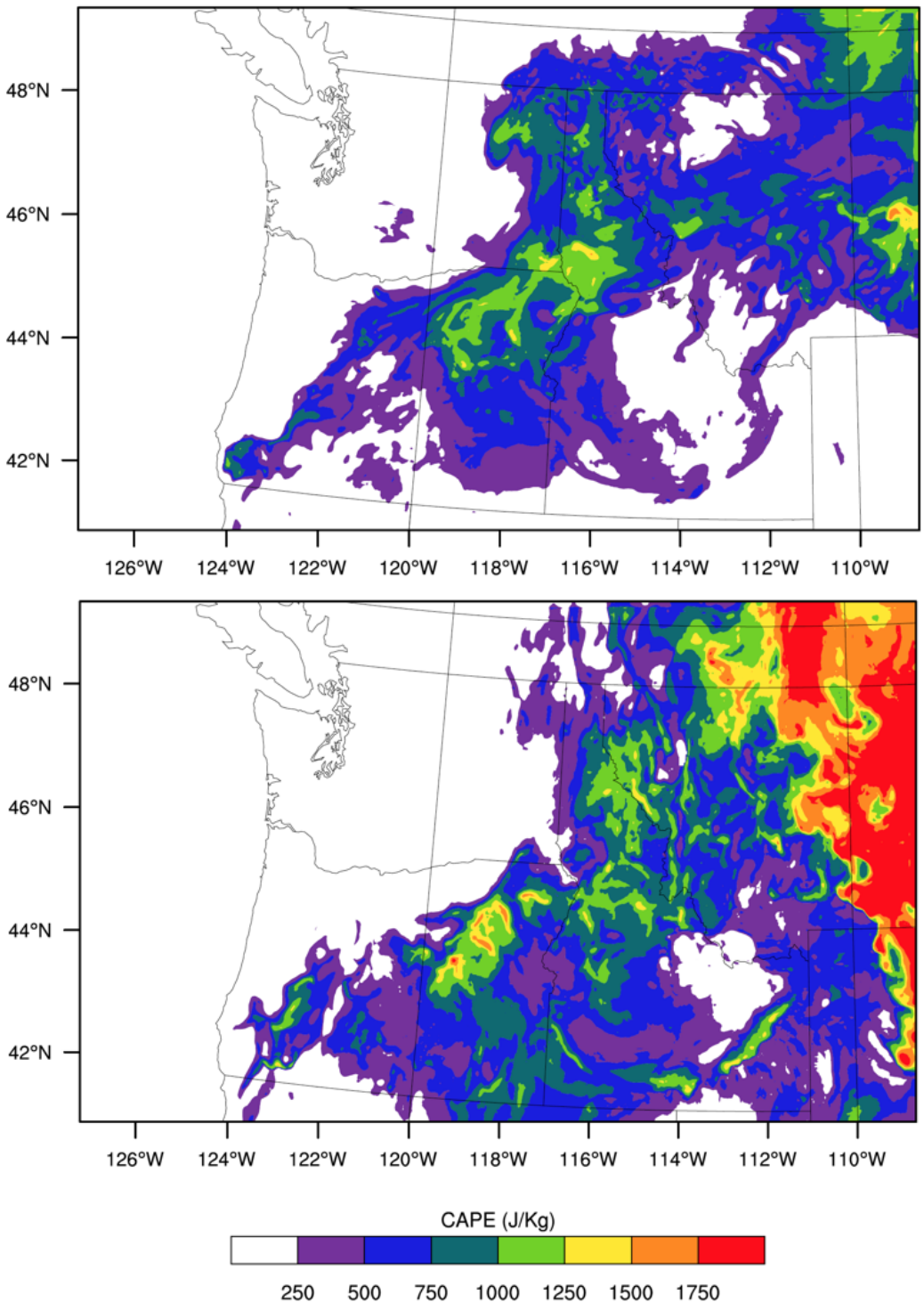


Figure 23. Pre-storm CAPE fields ($J\ kg^{-1}$) for the 1956 Meyers Canyon storm (top) and 1965 Lane Canyon storm (bottom).